



Sveučilište u Zagrebu

Fakultet strojarstva i brodogradnje

Marina Tošić

**MODELI ZA POTPORU PRI  
ODLUČIVANJU O RASPOLOŽIVOSTI  
ZRAKOPLOVA TEMELJEM DUBINSKE  
ANALIZE PODATAKA**

DOKTORSKI RAD

Zagreb, 2017.



University of Zagreb

Faculty of Mechanical Engineering and Naval Architecture

Marina Tošić

# **DECISION SUPPORT MODELS BASED ON DATA MINING TECHNIQUES FOR PREDICTING AIRCRAFT AVAILABILITY**

DOCTORAL THESIS

Zagreb, 2017.



Sveučilište u Zagrebu

Fakultet strojarstva i brodogradnje

Marina Tošić

**MODELI ZA POTPORU PRI  
ODLUČIVANJU O RASPOLOŽIVOSTI  
ZRAKOPLOVA TEMELJEM DUBINSKE  
ANALIZE PODATAKA**

DOKTORSKI RAD

Mentor: Izv. prof. dr. sc. Dragutin Lisjak

Zagreb, 2017.



University of Zagreb

Faculty of Mechanical Engineering and Naval Architecture

Marina Tošić

# **DECISION SUPPORT MODELS BASED ON DATA MINING TECHNIQUES FOR PREDICTING AIRCRAFT AVAILABILITY**

DOCTORAL THESIS

Supervisor: Dragutin Lisjak, PhD, Associate professor

Zagreb, 2017.

## Podaci za bibliografsku karticu

UDK:	629.7:004.62(043.3)
Ključne riječi:	raspoloživost zrakoplova, dubinska analiza podataka, klasifikacijski modeli, održavanje
Znanstveno područje:	Tehničke znanosti
Znanstveno polje:	Strojarstvo
Institucija:	Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje
Mentor:	Izv. prof. dr. sc. Dragutin Lisjak
Broj stranica:	191
Broj slika:	56
Broj tablica:	69
Broj korištenih bibliografskih jedinica:	105
Datum obrane:	27.10.2017.
Povjerenstvo:	Izv. prof. dr. sc. Milan Vrdoljak, predsjednik (Fakultet strojarstva i brodogradnje, Zagreb) Izv. prof. dr. sc. Dragutin Lisjak, mentor (Fakultet strojarstva i brodogradnje, Zagreb) Doc. dr. sc. Anita Domitrović, član (Fakultet prometnih znanosti, Zagreb)
Institucija u kojoj je rad pohranjen:	Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje Nacionalna i sveučilišna knjižnica u Zagrebu

## Zahvale

Zahvaljujem mentoru, izv. prof. dr. sc. Dragutinu Lisjaku, na oslobađanju od mnogobrojnih obaveza tijekom izrade ove disertacije, razumijevanju i podršci te na svim sugestijama koje su formirale ovo istraživanje.

Zahvaljujem članovima komisije, doc. dr. sc. Aniti Domitrović i izv. prof. dr. sc. Milanu Vrdoljaku, na izdvojenom vremenu za konzultacije i konstruktivnim savjetima u svim fazama izrade ove disertacije.

Hvala doc. dr. sc. Goranu Glavašu s Fakulteta poslovne informatike i matematike, Sveučilišta u Meinheimu i doc. dr. sc. Alanu Joviću s Fakulteta elektrotehnike i računarstva, Sveučilišta u Zagrebu, na inicijalnim savjetima oko pripreme podataka i izgradnje modela.

Hvala g. Robertu Mičugi na preporuci za ulazak u Croatia Airlines tvrtku.

Velika hvala g. Davoru Bujanu, dipl. ing. strojarstva, što mi je omogućio boravak u tvrtki i prikupljanje svih podataka bez kojih izrada ove disertacije ne bi bila moguća.

Hvala g. Silvestru Špeharu, dipl. ing. strojarstva, na korisnim diskusijama i ustupanju dozvola za ulazak u Croatia Airlines tvrtku.

Zahvaljujem g. Goranu Bjelici, dipl. ing. strojarstva, na pomoći oko interpretacije dijela podataka.

Zahvaljujem svim djelatnicima odjela Održavanje, a posebice g. Tomislavu Cerovskom, ing. elektrotehnike, na stručnim savjetima, evaluaciji rezultata te na nesebičnom prenošenju znanja.

Zahvaljujem prof. dr. sc. Nedeljku Štefaniću na povjerenju.

Hvala svim kolegama i prijateljima s Fakulteta strojarstva i brodogradnje na moralnoj potpori.

Borisu, Sandri, Zlatku, Mariji i Marku što ste me podržavali i uveseljavali na ovom putu.

Naposljetku, hvala mojim najmilijima – sestri Mirjani, mami Višnji i tati Dušku na neizmjernoj podršci tijekom školovanja.

Zagreb, rujan 2017.

Marina Tošić

## Sažetak

Proces održavanja zrakoplova karakteriziraju velike količine strukturiranih i nestrukturiranih podataka koji se svakodnevno bilježe u obliku pilotskih izvještaja, izvještaja o aktivnostima održavanja, zapisa o nezgodama i zastojsima, izvještaja nakon leta, itd. Znanja dobivena na temelju povijesnih podataka se mogu koristiti za unapređenje procesa održavanja zrakoplova, no ekstrakcija tih znanja se ne može obaviti ručno. U tu svrhu se koriste tehnike dubinske analize podataka, koje omogućavaju automatiziranu ili polu-automatiziranu ekstrakciju znanja iz skupova podataka.

Predloženo istraživanje bavi se razvojem modela, temeljenih na tehnikama dubinske analize podatka, koji će služiti kao potpora pri odlučivanju o raspoloživosti zrakoplova, a pri modeliranju se koriste tri izvora podataka: podaci prikupljeni iz sustava za nadzor tehničke ispravnosti stanja zrakoplova, zapisi o nepravilnostima u radu sustava/prošlim kvarovima i zapisi o zastojsima u eksploataciji zrakoplova (operacijskim zastojsima). Podaci su prikupljeni u razdoblju od 76 uzastopnih mjeseci za četiri zrakoplova tipa Airbus A319/20 koji pripadaju floti jednog zračnog prijevoznika. U disertaciji su analizirani kritični sustavi zrakoplova; sustav automatskog leta (ATA 22), sustav upravljanja letjelicom (ATA 27), hidraulički sustav (ATA 29), sustav podvozja (ATA 32) i navigacijski sustav (ATA 34).

Klasifikacijski modeli prve grupe izgrađeni su s ciljem predviđanja nastanka pilotskog upisa u tehničku knjigu na temelju prethodno generiranih poruka upozorenja u određenim fazama leta. Razvijen je postupak za strukturiranje podataka prikupljenih iz sustava za nadzor tehničke ispravnosti zrakoplova te postupak integracije tih podataka sa zapisima o nepravilnostima u radu sustava/prošlim kvarovima. Udruživanjem skupova podataka za izgradnju klasifikacijskih modela prve grupe otkriven je rang relevantnih značajki primjenom različitih filterskih postupaka (korelacijskog postupka, gini indeks postupka, informacijskog dobitka i omjera informacijskog dobitka) za selekciju značajki. Dodatno je provedeno istraživanje kako smanjenje udjela ulaznih značajki utječe na učinkovitost modela (*F-mjeru*, *osjetljivosti* i *specifičnosti*) za različite načine uzorkovanja podataka (uravnoteženo i slučajno uzorkovanje).

Klasifikacijski modeli druge grupe izgrađeni su s ciljem predviđanja posljedice nastalih tekstualnih pilotskih zapisa na raspoloživost zrakoplova. Izgradnji modela prethodila je aktivnost udruživanja skupova podataka o nepravilnostima u radu sustava/kvarova i

operacijskim zastojima. Usvajanjem evaluacijske mjere *točnosti*, modeli su uspoređeni s postojećim modelom iz sličnih istraživanja te je dokazana njihova primjenjivost.

*Ključne riječi:* raspoloživost zrakoplova, dubinska analiza podataka, klasifikacijski modeli, održavanje



## Extended summary

The aircraft maintenance process is characterized by large amounts of structured and unstructured data that are recorded daily in the form of pilot reports, maintenance logs, records of operational interruptions and technical incidents, post-flight reports, etc. The knowledge hidden within this data can potentially be used to improve the maintenance process, but its extraction can hardly be done manually. Therefore, in the last couple of years, a trend of development of predictive models using data mining techniques has been noticed. However, it can be concluded that these techniques are still not sufficiently applied in the process of aircraft maintenance because they require interdisciplinary knowledge, which includes understanding of the database, statistical knowledge, and understanding of machine learning and artificial intelligence techniques and models. This provides motivation for further research in this field.

The aim of the work presented in this thesis is to develop a new decision support models based on data mining techniques that will be used for predicting aircraft availability. In modelling process, three independent sources of data will be used; data collected from aircraft health monitoring system (AHMS), information of past faults/defects and information of operational interruptions. These data were collected over a period of 76 consecutive months for four Airbus A319/20 aircraft. Only data from critical aircraft systems were analysed in this dissertation; auto flight system (ATA 22), flight control system (ATA 27), hydraulic system (ATA 29), landing gear system (ATA 32) and navigation system (ATA 34). Based on the data collected from these systems, two groups of classification models were built.

The aim of the first group classification models is to determine whether a specific warning message/group of messages, collected from AHMS during different flight phases, will result in a pilot logbook entry. Prior to model development step, an algorithm for structuring AHMS data and algorithm for data fusion was developed. By integrating two data sources (warning messages from AHMS and information of past faults/defects), four filter methods (correlation based method, Gini index, information gain and information gain ratio) for feature selection were applied on a combined data source. In addition, research has been carried out to investigate how the reduction of the data dimensionality (a feature vector), in combination with different sampling techniques (stratified and shuffled), affects model performance measures (F-measure, sensitivity and specificity).

The aim of the second group classification models is to determine whether the created pilot logbook record will affect the aircraft availability, i.e. whether it will result in aircraft on ground (AOG) status, delay or flight cancellation status. This group of models is also built on combined data set, i.e. by integrating information of past faults/defects and information of operational interruptions. By adopting the evaluation measure of accuracy, developed models were compared to existing model from similar past research, and its applicability was demonstrated.

The research was carried out in several phases, which are summarized in the following chapters.

**Chapter 1** “Introduction” outlines literature gaps in the field of the aircraft maintenance. Based on the information acquired from published scientific papers and doctoral dissertations, various data mining techniques used for prediction in the field of aircraft maintenance were presented in the second subchapter. The rest of the sections present main aim of this research, hypothesis, expected scientific contributions, domain-specific terminology and an overview of the thesis.

**Chapter 2** “Prognostics in the aircraft maintenance process” highlights the benefits and challenges of the prognostics approaches currently applied in the literature. It explains in detail knowledge data discovery process, as well as supervised and unsupervised data mining techniques. Cluster analysis, association rules and various classification algorithms (Neural Networks, Decision Trees, Support Vector Machines and Naïve Bayes) were additionally described. Four filtering methods for feature selection were introduced. Different data sampling techniques and measures for evaluation of classification models were also described in this chapter.

**Chapter 3** “Processing and analysis of collected data” provides a detail explanation of the three independent data sources (AHMS data, information of past faults/defects and information of operational interruptions) used for this research. This chapter also presents a procedure for structuring AHMS data based on algorithm presented in Appendix A. Except for this procedure, two additional procedures for data fusion are shown; a procedure for data fusion of AHMS data and past faults/defects (Appendix B), and procedure for data fusion of past faults/defects and data containing information of operational interruptions. The final result of this procedures presents two combined data sets, which are transformed into a form suitable for modelling. In addition, an exploratory data analysis was conducted on these combined data sets and recommended guidelines were provided for future research phases.

**Chapter 4** “The process of discovering relevant features and the development of classification models” consists of the concise research steps used for building various data mining models. In the first section, an outline of the process developed within RapidMiner platform for discovering relevant features by applying filtering methods is given. This section also presents a process for applying the association rule mining to the combined data set to identify the features that appear more frequently together. The second section presents a process for building a first group classification models, while the third one presents a process for building a second group classification models. A list of the operators and their parameters used for building the models is presented. A graphical representation of these models is given in Appendix C.

**Chapter 5** “Research results” outlines the results of the models developed in previous chapters. The resulting models were tested on a verification data set, i.e. a data set that was not used for model building. To determine the performance of the built classification models, evaluation metric has been used. Due to the data imbalance, the first group classification models were evaluated by F-measure, sensitivity and specificity. All models were first built on the original dataset, which contains all the features. During the modelling phase, different data sampling techniques (stratified and shuffled) were used. By applying different filter methods, a rank of the features, representing a feature vector, was obtained (Appendix D). In order to evaluate results of the filter methods, the performance of the first group classification models was observed by gradually decreasing the number of the features from the rank obtained by each filter method. The main metric for the performance evaluation of the second group classification models was accuracy. After obtaining model accuracy, these results were compared with results from the similar model found in the literature.

**Chapter 6** “Conclusion” presents the final chapter where the original scientific contributions are presented and a summary of the results is provided. Also, theoretical and practical implications are outlined and areas for potential future research are suggested.

*Keywords:* aircraft availability, data mining, classification models, maintenance

# Sadržaj

Popis slika .....	V
Popis tablica .....	VII
Popis oznaka i kratica.....	X
1. UVOD .....	1
1.1. Uvodna razmatranja.....	1
1.2. Motivacija i pregled dosadašnjih istraživanja .....	3
1.3. Cilj istraživanja, hipoteza rada i očekivani znanstveni doprinosi .....	9
1.4. Terminologija područja .....	10
1.5. Struktura rada .....	11
2. PROGNOZIRANJE U PROCESU ODRŽAVANJA ZRAKOPLOVA .....	13
2.1. Uvodna razmatranja.....	13
2.2. Podjela prognostičkih pristupa .....	14
2.3. Dubinska analiza podataka .....	16
2.3.1. Proces otkrivanja znanja u bazama podataka .....	16
2.3.2. Nenadzirano učenje .....	17
2.3.2.1. Klaster analiza.....	17
2.3.2.2. Asocijativna pravila .....	18
2.3.3. Nadzirano učenje.....	19
2.3.3.1. Neuronske mreže .....	19
2.3.3.2. Stablo odlučivanja.....	21
2.3.3.3. Metoda potpornih vektora.....	22
2.3.3.4. Naive Bayes .....	24
2.3.4. Postupci selekcije značajki.....	28
2.3.4.1. Informacijski dobitak .....	29
2.3.4.2. Omjer informacijskog dobitka .....	30
2.3.4.3. Gini indeks .....	30

2.3.4.4.	Korelacija .....	31
2.3.5.	Evalvacija i optimizacija klasifikacijskih modela .....	32
2.3.5.1.	Tehnike probira za vrednovanje performansi modela .....	32
2.3.5.2.	Mjere za vrednovanje performansi modela.....	33
2.3.5.3.	Optimizacija modela .....	35
2.3.6.	Alati za dubinsku analizu podataka.....	36
2.4.	Zaključak poglavlja i smjernice za nastavak istraživanja.....	37
3.	OBRADA I ANALIZA PRIKUPLJENIH PODATAKA .....	39
3.1.	Uvodna razmatranja.....	39
3.2.	Opis dostupnih podataka .....	40
3.2.1.	Podaci prikupljeni iz AHMS-a.....	41
3.2.2.	Podaci o prošlim kvarovima ili nepravilnostima u radu sustava.....	44
3.2.3.	Podaci o zastoјima u eksploataciji zrakoplova.....	46
3.3.	Strukturiranje i udruživanje skupova podataka .....	48
3.3.1.	Postupak strukturiranja i integracija skupa WN i PI-DE .....	48
3.3.2.	Postupak integracije skupa PI-DE i skupa OZ .....	54
3.4.	Eksplorativna analiza podataka .....	55
3.4.1.	Skup WN i DE zapisa.....	55
3.4.2.	Skup DE i OZ zapisa.....	59
3.5.	Zaključak poglavlja i smjernice za nastavak istraživanja.....	60
4.	FORMULACIJA PROCESA ZA OTKRIVANJE RELEVANTNIH ZNAČAJKI I IZGRADNJU KLASIFIKACIJSKIH MODELA .....	62
4.1.	Formulacija procesa za otkrivanje relevantnih značajki.....	62
4.1.1.	Priprema skupa podataka .....	64
4.1.2.	Otkrivanje relevantnih značajki .....	64
4.1.3.	Otkrivanje asocijacija unutar značajki .....	67
4.2.	Formulacija procesa za izgradnju klasifikacijskih modela prve grupe.....	68

4.2.1.	Priprema skupa podataka .....	69
4.2.2.	Izgradnja modela .....	71
4.3.	Formulacija procesa za izgradnju klasifikacijskih modela druge grupe.....	73
4.3.1.	Priprema skupa podataka .....	75
4.3.2.	Izgradnja modela .....	77
4.4.	Zaključak poglavlja i smjernice za nastavak istraživanja.....	78
5.	PRIKAZ I EVALUACIJA REZULTATA ISTRAŽIVANJA .....	79
5.1.	Uvodna razmatranja.....	79
5.2.	Rezultati ranga relevantnih značajki i asocijativnih pravila.....	80
5.2.1.	Rezultati ranga relevantnih značajki .....	80
5.2.1.1.	ATA 22 .....	81
5.2.1.2.	ATA 27 .....	82
5.2.1.3.	ATA 29 .....	83
5.2.1.4.	ATA 32 .....	85
5.2.1.5.	ATA 34 .....	86
5.2.2.	Rezultati asocijativnih pravila.....	87
5.3.	Rezultati klasifikacijskih modela prve grupe .....	88
5.3.1.	ATA 22.....	88
5.3.1.1.	Prikaz i evaluacija rezultata modela za zrakoplov $Ac_1$ .....	89
5.3.1.2.	Prikaz i evaluacija rezultata modela za zrakoplov $Ac_2$ .....	92
5.3.1.3.	Prikaz i evaluacija rezultata modela za zrakoplov $Ac_3$ .....	95
5.3.1.4.	Prikaz i evaluacija rezultata modela za zrakoplov $Ac_4$ .....	98
5.3.2.	ATA 27.....	101
5.3.2.1.	Prikaz i evaluacija rezultata modela za zrakoplov $Ac_1$ .....	101
5.3.2.2.	Prikaz i evaluacija rezultata modela za zrakoplov $Ac_2$ .....	104
5.3.2.3.	Prikaz i evaluacija rezultata modela za zrakoplov $Ac_3$ .....	107
5.3.2.4.	Prikaz i evaluacija rezultata modela za zrakoplov $Ac_4$ .....	110

5.3.3.	ATA 32.....	113
5.3.3.1.	Prikaz i evaluacija rezultata modela za zrakoplov Ac <sub>1</sub> .....	113
5.3.3.2.	Prikaz i evaluacija rezultata modela za zrakoplov Ac <sub>2</sub> .....	116
5.3.3.3.	Prikaz i evaluacija rezultata modela za zrakoplov Ac <sub>3</sub> .....	119
5.3.3.4.	Prikaz i evaluacija rezultata modela za zrakoplov Ac <sub>4</sub> .....	122
5.3.4.	ATA 34.....	125
5.3.4.1.	Prikaz i evaluacija rezultata modela za zrakoplov Ac <sub>1</sub> .....	125
5.3.4.2.	Prikaz i evaluacija rezultata modela za zrakoplov Ac <sub>2</sub> .....	128
5.3.4.3.	Prikaz i evaluacija rezultata modela za zrakoplov Ac <sub>3</sub> .....	131
5.3.4.4.	Prikaz i evaluacija rezultata modela za zrakoplov Ac <sub>4</sub> .....	134
5.4.	Rezultati klasifikacijskih modela druge grupe .....	137
5.4.1.	ATA 22.....	137
5.4.2.	ATA 27.....	139
5.4.3.	ATA 29.....	140
5.4.4.	ATA 32.....	142
5.4.5.	ATA 34.....	143
5.5.	Zaključak poglavlja .....	145
6.	ZAKLJUČAK .....	148
6.1.	Osvrt na početna istraživanja.....	148
6.2.	Osvrt na znanstvene doprinose i hipotezu istraživanja.....	148
6.3.	Teorijske i praktične implikacije .....	150
6.4.	Ograničenja istraživanja .....	151
6.5.	Smjernice za daljnja istraživanja .....	152
	LITERATURA.....	154
	PRILOZI.....	165
	Životopis.....	190
	Curriculum vitae.....	191

## Popis slika

Slika 2-1. <i>Arhitektura neurona s <math>X_i</math> ulaznih značajki</i> [54] .....	19
Slika 2-2. <i>Arhitektura neuronskih mreža</i> [54].....	20
Slika 2-3. <i>Prikaz linearnog razdvajanja prostora ulaznih podataka</i> [54] .....	22
Slika 2-4. <i>Prikaz nelinearnog razdvajanja prostora ulaznih podataka</i> [54] .....	23
Slika 2-5. <i>Prikaz svrstavanja ulaznih uzoraka u prostor značajki gdje su razredi linearno razdjeljivi</i> [54] .....	23
Slika 2-6. <i>Prikaz uzoraka podijeljenih u dvije klase</i> [54] .....	25
Slika 2-7. <i>Pozicija novog uzorka za koji je potrebno odrediti klasu</i> [54].....	26
Slika 2-8. <i>Usporedba metoda procjene gustoće vjerojatnosti numeričkih značajki</i> [59] .....	27
Slika 3-1. <i>Prikaz izvornog zapisa poruka upozorenja prikupljenih s PFR-a</i> .....	42
Slika 3-2. <i>Prikaz logičkih cjelina jednog zapisa poruke upozorenja</i> .....	42
Slika 3-3. <i>Prikaz faza leta zrakoplova</i> [95].....	44
Slika 3-4. <i>Dijagram toka postupka za strukturiranje i udruživanje skupova WN i PI-DE</i> .....	49
Slika 3-5. <i>Prikaz značajki skupa <math>D_{WN}</math></i> .....	50
Slika 3-6. <i>Prikaz značajki skupa <math>D_{DE}</math></i> .....	51
Slika 3-7. <i>Prikaz dodjeljivanja šifre svakoj poruci upozorenja unutar skupa <math>D_{WN}</math></i> .....	51
Slika 3-8. <i>Prikaz stvaranja hibridne ulazne značajke unutar skupa <math>D_{WN}</math></i> .....	52
Slika 3-9. <i>Prikaz transformacije skupa <math>D_{WN}</math> u skup <math>D_{WN}^T</math> primjenom modula Pivot Table</i> ...	52
Slika 3-10. <i>Prikaz preprocesuiranja skupa <math>D_{WN}^T</math></i> .....	53
Slika 3-11. <i>Prikaz skupa podataka <math>D_{KMI}</math></i> .....	53
Slika 3-12. <i>Dijagram toka postupaka za udruživanje skupova PI-DE i OZ</i> .....	54
Slika 3-13. <i>Prikaz dijela uzoraka iz skupa podataka <math>D_{KM2}</math> zrakoplova <math>Ac_i</math></i> .....	55
Slika 3-14. <i>Prikaz broja poruka upozorenja po analiziranim sustavima na razini flote</i> .....	56
Slika 3-15. <i>Prikaz frekvencije pojava poruka upozorenja po fazama leta za ATA 22, ATA 27, ATA 29, ATA 32 i ATA 34 sustave</i> .....	56
Slika 3-16. <i>Usporedba broja generiranih WN poruka i PI-DE zapisa nastalih temeljem istih poruka po svakom sustavu</i> .....	57
Slika 3-17. <i>Prikaz broja uzoraka u skupu <math>D_{KMI}</math> za <math>Ac_i</math> zrakoplov i analizirani sustav</i> .....	58
Slika 3-18. <i>Klasifikacija operacijskih zastoja na razini analiziranih sustava</i> .....	59
Slika 3-19. <i>Usporedba broja PI-DE zapisa i OZ zapisa po analiziranim sustavima</i> .....	60



Slika 4-1. <i>Dijagram toka procesa za otkrivanje relevantnih značajki i asocijacija unutar <math>D_{KMIN}</math> skupova podataka</i> .....	63
Slika 4-2. <i>Prikaz skupa podataka <math>D_{KMIN}</math> za sustav <math>S_i</math></i> .....	64
Slika 4-3. <i>Grafički prikaz procesa za otkrivanje ranga značajki</i> .....	66
Slika 4-4. <i>Dijagram toka procesa izgradnje klasifikacijskih modela prve grupe</i> .....	69
Slika 4-5. <i>Prikaz skupa podataka <math>D_{KM1-1}</math></i> .....	70
Slika 4-6. <i>Prikaz izlaznih rezultata iz klasifikacijskog modela prve grupe</i> .....	73
Slika 4-7. <i>Dijagram toka procesa izgradnje klasifikacijskih modela druge grupe</i> .....	74
Slika 4-8. <i>Prikaz stvaranja skupa <math>D_{KM2T}</math></i> .....	76
Slika 5-1. <i>ATA 22 – prikaz ranga prvih 50% značajki</i> .....	82
Slika 5-2. <i>ATA 27 – prikaz ranga prvih 50% značajki</i> .....	83
Slika 5-3. <i>ATA 29 – prikaz ranga prvih 50% značajki</i> .....	84
Slika 5-4. <i>ATA 32 – prikaz ranga prvih 50% značajki</i> .....	86
Slika 5-5. <i>ATA 34 – prikaz ranga prvih 50% značajki</i> .....	87
Slika 5-6. <i>ATA 22 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_1</math></i> .....	90
Slika 5-7. <i>ATA 22 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_2</math></i> .....	93
Slika 5-8. <i>ATA 22 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_3</math></i> .....	96
Slika 5-9. <i>ATA 22 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_4</math></i> .....	99
Slika 5-10. <i>ATA 27 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_1</math></i> .....	102
Slika 5-11. <i>ATA 27 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_2</math></i> .....	105
Slika 5-12. <i>ATA 27 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_3</math></i> .....	108
Slika 5-13. <i>ATA 27 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_4</math></i> .....	111
Slika 5-14. <i>ATA 32 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_1</math></i> .....	114
Slika 5-15. <i>ATA 32 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_2</math></i> .....	117
Slika 5-16. <i>ATA 32 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_3</math></i> .....	120
Slika 5-17. <i>ATA 32 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_4</math></i> .....	123
Slika 5-18. <i>ATA 34 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_1</math></i> .....	126
Slika 5-19. <i>ATA 34 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_2</math></i> .....	129
Slika 5-20. <i>ATA 34 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_3</math></i> .....	132
Slika 5-21. <i>ATA 34 – prikaz vrijednosti <math>F</math>-mjere za <math>Ac_4</math></i> .....	135

## Popis tablica

Tablica 2-1. <i>Prikaz matrice zabune za kategoričku klasifikaciju [68]</i> .....	33
Tablica 2-2. <i>Prikaz evaluacijskih mjera temeljenih na analizi matrice zabune [68]</i> .....	34
Tablica 3-1. <i>Popis dijela tehničkih karakteristika zrakoplova</i> .....	41
Tablica 3-2. <i>Objašnjenje strukture poruke upozorenja</i> .....	43
Tablica 3-3. <i>Prikaz značajki koje opisuju zapis o nepravilnosti/kvaru</i> .....	45
Tablica 3-4. <i>Prikaz značajki koje opisuju podatke o zastojsima u eksploataciji zrakoplova</i> ....	47
Tablica 3-5. <i>Pseudo kôd algoritma za strukturiranje WN poruka</i> .....	50
Tablica 4-1. <i>Proces otkrivanja relevantnih značajki</i> .....	65
Tablica 4-2. <i>Proces otkrivanja asocijativnih pravila</i> .....	67
Tablica 4-3. <i>Prikaz procesa za stvaranje <math>D_{KMI-1}</math> skupa podataka</i> .....	70
Tablica 5-1. <i>Asocijativna pravila za značajku „06/29-12“</i> .....	88
Tablica 5-2. <i>ATA 22 zrakoplov <math>Ac_1</math> - vrijednosti F-mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka</i> .....	89
Tablica 5-3. <i>ATA 22 – prikaz vrijednosti mjere osjetljivosti za <math>Ac_1</math></i> .....	91
Tablica 5-4. <i>ATA 22 – prikaz vrijednosti mjere specifičnosti za <math>Ac_1</math></i> .....	91
Tablica 5-5. <i>ATA 22 zrakoplov <math>Ac_2</math> - vrijednosti F-mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka</i> .....	92
Tablica 5-6. <i>ATA 22 – prikaz vrijednosti mjere osjetljivosti za <math>Ac_2</math></i> .....	94
Tablica 5-7. <i>ATA 22 – prikaz vrijednosti mjere specifičnosti za <math>Ac_2</math></i> .....	94
Tablica 5-8. <i>ATA 22 zrakoplov <math>Ac_3</math> - vrijednosti F-mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka</i> .....	95
Tablica 5-9. <i>ATA 22 – prikaz vrijednosti mjere osjetljivosti za <math>Ac_3</math></i> .....	97
Tablica 5-10. <i>ATA 22 – prikaz vrijednosti mjere specifičnosti za <math>Ac_3</math></i> .....	97
Tablica 5-11. <i>ATA 22 zrakoplov <math>Ac_4</math> - vrijednosti F-mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka</i> .....	98
Tablica 5-12. <i>ATA 22 – prikaz vrijednosti mjere osjetljivosti za <math>Ac_4</math></i> .....	100
Tablica 5-13. <i>ATA 22 – prikaz vrijednosti mjere specifičnosti za <math>Ac_4</math></i> .....	100
Tablica 5-14. <i>ATA 27 zrakoplov <math>Ac_1</math> - vrijednosti F-mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka</i> .....	101
Tablica 5-15. <i>ATA 27 – prikaz vrijednosti mjere osjetljivosti za <math>Ac_1</math></i> .....	103
Tablica 5-16. <i>ATA 27 – prikaz vrijednosti mjere specifičnosti za <math>Ac_1</math></i> .....	103

Tablica 5-17. ATA 27 zrakoplov $Ac_2$ - vrijednosti $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka.....	104
Tablica 5-18. ATA 27 – prikaz vrijednosti mjere osjetljivosti za $Ac_2$ .....	106
Tablica 5-19. ATA 27 – prikaz vrijednosti mjere specifičnosti za $Ac_2$ .....	106
Tablica 5-20. ATA 27 zrakoplov $Ac_3$ - vrijednosti $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka.....	107
Tablica 5-21. ATA 27 – prikaz vrijednosti mjere osjetljivosti za $Ac_3$ .....	109
Tablica 5-22. ATA 27 – prikaz vrijednosti mjere specifičnosti za $Ac_3$ .....	109
Tablica 5-23. ATA 27 zrakoplov $Ac_4$ - vrijednosti $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka.....	110
Tablica 5-24. ATA 27 – prikaz vrijednosti mjere osjetljivosti za $Ac_4$ .....	112
Tablica 5-25. ATA 27 – prikaz vrijednosti mjere specifičnosti za $Ac_4$ .....	112
Tablica 5-26. ATA 32 zrakoplov $Ac_1$ - vrijednosti $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka.....	113
Tablica 5-27. ATA 32 – prikaz vrijednosti mjere osjetljivosti za $Ac_1$ .....	115
Tablica 5-28. ATA 32 – prikaz vrijednosti mjere specifičnosti za $Ac_1$ .....	115
Tablica 5-29. ATA 32 zrakoplov $Ac_2$ - vrijednosti $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka.....	116
Tablica 5-30. ATA 32 – prikaz vrijednosti mjere osjetljivosti za $Ac_2$ .....	118
Tablica 5-31. ATA 32 – prikaz vrijednosti mjere specifičnosti za $Ac_2$ .....	118
Tablica 5-32. ATA 32 zrakoplov $Ac_3$ - vrijednosti $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka.....	119
Tablica 5-33. ATA 32 – prikaz vrijednosti mjere osjetljivosti za $Ac_3$ .....	121
Tablica 5-34. ATA 32 – prikaz vrijednosti mjere specifičnosti za $Ac_3$ .....	121
Tablica 5-35. ATA 32 zrakoplov $Ac_4$ - vrijednosti $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka.....	122
Tablica 5-36. ATA 32 – prikaz vrijednosti mjere osjetljivosti za $Ac_4$ .....	124
Tablica 5-37. ATA 32 – prikaz vrijednosti mjere specifičnosti za $Ac_4$ .....	124
Tablica 5-38. ATA 34 zrakoplov $Ac_1$ - vrijednosti $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka.....	125
Tablica 5-39. ATA 34 – prikaz vrijednosti mjere osjetljivosti za $Ac_1$ .....	127
Tablica 5-40. ATA 34 – prikaz vrijednosti mjere specifičnosti za $Ac_1$ .....	127
Tablica 5-41. ATA 34 zrakoplov $Ac_2$ - vrijednosti $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka.....	128

Tablica 5-42. ATA 34 – prikaz vrijednosti mjere osjetljivosti za $Ac_2$ .....	130
Tablica 5-43. ATA 34 – prikaz vrijednosti mjere specifičnosti za $Ac_2$ .....	130
Tablica 5-44. ATA 34 zrakoplov $Ac_3$ - vrijednosti F-mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka.....	131
Tablica 5-45. ATA 34 – prikaz vrijednosti mjere osjetljivosti za $Ac_3$ .....	133
Tablica 5-46. ATA 34 – prikaz vrijednosti mjere specifičnosti za $Ac_3$ .....	133
Tablica 5-47. ATA 34 zrakoplov $Ac_4$ - vrijednosti F-mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka.....	134
Tablica 5-48. ATA 34 – prikaz vrijednosti mjere osjetljivosti za $Ac_4$ .....	136
Tablica 5-49. ATA 34 – prikaz vrijednosti mjere specifičnosti za $Ac_4$ .....	136
Tablica 5-50. ATA 22 – prikaz matrice zabune kada je zavisna značajka Utjecaj na raspoloživost.....	137
Tablica 5-51. ATA 22 – prikaz matrice zabune kada je zavisna značajka Posljedica .....	138
Tablica 5-52. ATA 27 – prikaz matrice zabune kada je zavisna značajka Utjecaj na raspoloživost.....	139
Tablica 5-53. ATA 27 – prikaz matrice zabune kada je zavisna značajka Posljedica .....	140
Tablica 5-54. ATA 29 – prikaz matrice zabune kada je zavisna značajka Utjecaj na raspoloživost.....	141
Tablica 5-55. ATA 29 – prikaz matrice zabune kada je zavisna značajka Posljedica .....	141
Tablica 5-56. ATA 32 – prikaz matrice zabune kada je zavisna značajka Utjecaj na raspoloživost.....	142
Tablica 5-57. ATA 32 – prikaz matrice zabune kada je zavisna značajka Posljedica .....	143
Tablica 5-58. ATA 34 – prikaz matrice zabune kada je zavisna značajka Utjecaj na raspoloživost.....	144
Tablica 5-59. ATA 34 – prikaz matrice zabune kada je zavisna značajka Posljedica .....	144

## Popis oznaka i kratica

Popis oznaka		
Oznaka	Značenje	Mjerna jedinica
$D$	Skup podataka	—
$d$	Broj uzoraka u skupu podataka	—
$d_i$	Broj uzoraka u pojedinoj klasi	—
$X_i$	Vektor značajki	—
$A_i$	Ulazna značajka	—
$x_i$	Vrijednost značajke	—
$C_i$	Klasa (razred)	—
$Ac_i$	Zrakoplov	—
$S_i$	Sustav zrakoplova	—
$D_{WN}$	Skup podataka poruka upozorenja	—
$D_{WN}^T$	Transformirani skup podataka poruka upozorenja	—
$D_{DE}$	Skup podataka o kvarovima/nepravilnostima u radu sustava	—
$D_{KM1}$	Skup podataka za izradu klasifikacijskih modela prve grupe	—
$D_{OZ}$	Skup podataka o zastojima u eksploataciji zrakoplova (operacijskim zastojima)	—
$D_{KM2}$	Skup podataka za izradu klasifikacijskih modela druge grupe	—
$R_{KMI}$	Rang značajki	—

## Popis kratica

Kratica	Značenje
ATA	Udruženje zračnih prijevoznika Amerike (eng. <i>Air Transports Association of America</i> )
WN	Poruke upozorenja (eng. <i>warning messages</i> )
PFR	Izvještaji nakon leta (eng. <i>post-flight report</i> )
ECAM	Centralizirani elektronični sustav (eng. <i>electronic centralised aircraft monitor</i> )
TLB	Tehnička knjiga zrakoplova (eng. <i>technical logbook</i> )
MEL	Lista minimalne ispravnosti opreme (eng. <i>minimum equipment list</i> )
MRO	Održavanje, popravak i remont (eng. <i>maintenance, repair and overhaul</i> )
CBM	Održavanje po stanju (eng. <i>Condition Based Maintenance</i> )
PdM	Prediktivno održavanje (eng. <i>Predictive Maintenance</i> )
IATA	Međunarodno udruženje zračnih prijevoznika (eng. <i>International Air Transport Association</i> )
AHMS	Sustav za nadzor tehničke ispravnosti zrakoplova (eng. <i>aircraft health monitoring system</i> )
NB	Naivni Bayes (eng. <i>Naive Bayes</i> )
NN	Neuronska mreža (eng. <i>Neural network</i> )
DT	Stablo odlučivanja (eng. <i>decision trees</i> )
PCA	Analiza glavnih komponenata (eng. <i>principal component analysis</i> )
CDA	Kanonička diskriminativna analiza (eng. <i>canonical discriminantal analysis</i> )
PSO	Algoritam optimizacija roja čestica (eng. <i>particle swarm optimization</i> )

---

### Popis kratica

---

Kratica	Značenje
SVM	Algoritam temeljen na potpornim vektorima (eng. <i>support vector machines</i> )
ACO	Optimizacija kolonijom mrava (eng. <i>ant colony optimization</i> )
<i>k</i> -NN	Algoritam <i>k</i> -najbližih susjeda (eng. <i>k-nearest neighbour</i> )
RT	Regresijska stabla (eng. <i>regression trees</i> )
KDD	Proces otkrivanja znanja u bazama podataka (eng. <i>Knowledge Discovery in Databases</i> )
IFAC	Međunarodno udruženje za automatsko upravljanje (eng. <i>International Federation of Automatic Control</i> )
PHM	Arhitektura za upravljanje i prognozu stanja tehničke ispravnosti zrakoplova (eng. <i>Prognostics and Health Management</i> )
FDR	Snimač podataka leta (eng. <i>Flight Data Recorder</i> )
GI	Gini indeks (eng. <i>gini index</i> )
IG	Informacijski dobitak (eng. <i>information gain</i> )
IGR	Omjer informacijskog dobitka (eng. <i>information gain ratio</i> )
KP	Korelacijski postupak (eng. <i>correlation</i> )
RM	RapidMiner alat
DE	Kvar/nepravilnost u radu sustava (eng. <i>defect</i> )
OZ	Operacijski zastoj
PI-DE	Zapis o prošlim kvarovima/nepravilnostima u radu sustava nastao od strane pilota
MA-DE	Zapis o prošlim kvarovima/nepravilnostima u radu sustava nastao od tehničkog osoblja

---

### Popis kratica

---

Kratica	Značenje
QA-DE	Zapis o prošlim kvarovima/nepravilnostima u radu sustava nastao temeljem zahtjeva odjela Osiguranja kvalitete i sigurnosti
A/C	Zrakoplov (eng. <i>aircraft</i> )
Acreg	Registracija zrakoplova (eng. <i>aircraft registration</i> )
Actype	Tip zrakoplova (eng. <i>aircraft type</i> )
Acc.Date	Datum kada je nepravilnost/kvar otklonjen
EASA	Europska agencija za sigurnost zračnog prometa (eng. <i>European Aviation Safety Agency</i> )
AMP	Program održavanja zrakoplova (eng. <i>Aircraft Maintenance Program</i> )
AOG	Prizemljenje zrakoplova (eng. <i>aircraft on ground</i> )
DTTRR	Odgođeno vrijeme popravka kvara (eng. <i>delay time to repair</i> )
STR	Uravnoteženi načina uzorkovanja podataka (eng. <i>stratified</i> )
SHU	Slučajni načina uzorkovanja podataka (eng. <i>shuffled</i> )
ver	Verifikacijski skup podataka (eng. <i>verification</i> )

---



# 1. UVOD

U uvodnom dijelu rada objašnjeni su problemi područja koji uključuju analizu podataka iz različitih izvora u procesu održavanja zrakoplova. Dan je pregled prošlih istraživanja kako bi se objasnila motivacija za istraživanje u ovoj disertaciji. Postavljen je cilj rada i hipoteza rada s očekivanim znanstvenim doprinosima. Objašnjena je domenska terminologija i prikazana je struktura disertacije.

## 1.1. Uvodna razmatranja

Zrakoplovni sustavi kategorizirani su u ATA (eng. (i dalje) *Air Transports Association of America*, ATA) poglavlja koja obuhvaćaju unificirane kôdove za grupaciju općih zahvata<sup>1</sup> na zrakoplovu (eng. *aircraft general*, ATA 00 – ATA 18), sustave i opremu zrakoplova (eng. *aircraft systems*, ATA 20 - ATA 50), strukturu zrakoplova (eng. *aircraft structure*, ATA 51 - ATA 57), sustava propelera/rotora (eng. *propeller/rotor*, ATA 60 - ATA 67) i pogonskih sustava (eng. *power plant*, ATA 70 - ATA 92). Drugim riječima, ATA kôd predstavlja određeni sustav (zapis od dvije znamenke), podsustav (zapis od četiri znamenke) ili komponentu (zapis od osam znamenaka).

Sa svrhom praćenja i nadzora ispravnog funkcioniranja sustava, zrakoplovi su opremljeni računalima koja generiraju predefinirane poruke upozorenja (eng. *warning messages*, WN) u različitim fazama leta. Neke poruke upozorenja mogu ukazivati na nepravilnosti u radu sustava te su bitne za rješavanje kako ne bi nastao zastoј sustava, dok druge poruke upozorenja mogu biti „lažne“ (eng. *spurious*). Ipak, sve poruke upozorenja bilježe se na izvještajima nakon leta (eng. (i dalje) *post-flight report*, PFR), a većina poruka isporučuje se centraliziranom elektroničnom sustavu (eng. (i dalje) *electronic centralised aircraft monitor*, ECAM) kako bi se dala indikacija pilotima o potencijalnoj nepravilnosti u radu sustava. Dugim riječima, ukoliko računalo registrira određenu nepravilnost u radu sustava, ECAM će upozoriti pilote s određenom porukom upozorenja i osigurati elektronički postupak koji može pomoći u rješavanju nepravilnosti.

---

<sup>1</sup> Opći zahvati na zrakoplovu podrazumijevaju periodične zahvate održavanja/provjere održavanja, izravnavanje i vaganje zrakoplova, vuču i taksiranje, plakate i oznake, itd.

U ECAM-u se poruke upozorenja mogu svrstati u tri kategorije - klase 1 (eng. *class 1*), klase 2 (eng. *class 2*) i klase 3 (eng. *class 3*). Za razliku od prethodne dvije kategorije, treća kategorija poruka nije vidljiva pilotima tokom leta na ECAM zaslonu i ne bilježi se na PFR-u.

Od navedenih kategorija, poruke klase 1 u većini slučajeva imaju za posljedicu operativni zastoj. Ukoliko dođe do pojave poruke ove klase tijekom vremena kada je zrakoplov u funkciji, pilot će unijeti u tehničku knjigu zrakoplova (eng. (i dalje) *technical logbook*, TLB) zapis, kako bi tehničko osoblje (eng. *maintenance crew*) provjerilo ispravnost sustava za koji se poruka pojavila. Svaki zapis u TLB, tj. pilotski upis nastao temeljem određene poruke upozorenja, može utjecati na raspoloživost zrakoplova, sukladno listi minimalne ispravnosti opreme<sup>2</sup> (eng. *minimum equipment list*, MEL).

S druge strane, poruke upozorenja klase 2 obično nisu povezane s kritičnim nepravilnostima u radu sustava koje bi mogle uzrokovati operacijske zastoje. Bitno je napomenuti da poruke klase 2 mogu prerasti u poruke klase 1 te je potrebno reagirati i na te poruke i ispraviti nepravilnosti u radu sustava koje uzrokuju pojavu tih poruka.

Stoga je važno analizirati poruke upozorenja jer njihov unos u TLB zahtijeva određene neplanirane aktivnosti održavanja (testiranje, popravak ili zamjenu komponente). Budući da neplanirane aktivnosti zahtijevaju korektivne zahvate održavanja, ti zahvati nisu obuhvaćeni inicijalnom planom održavanja. Također, sve neplanirane aktivnosti održavanja zahtijevaju dostupnost materijalnih i ljudskih resursa u kratkom vremenskom razdoblju, što utječe na raspoloživost zrakoplova. Samim time, analiza tekstualnih zapisa iz TLB-a i zapisa koji su uzrokovali operacijske zastoje može osigurati dostupnost materijalnih resursa.

Međutim, veza između poruka upozorenja i zapisa o nepravilnostima nastalim temeljem tih poruka, kao i zapisa koji su uzrokovali operacijske zastoje, nije izravna. Ovaj problem se javlja iz niza razloga, a neki od njih su to što zapisi u TLB nastali temeljem poruka upozorenja nemaju posebnu oznaku te dio pilotskih zapisa može proizaći i iz mehaničkih oštećenja. Također, ista poruka upozorenja može biti povezana s nepravilnosti u radu različitih komponenti što komplicira postupak identifikacije grešaka (eng. *troubleshooting*) koji se provodi iterativnim praćenjem procedura. Nadalje, zbog velike količine podataka koji se svakodnevno bilježe u

---

<sup>2</sup> MEL sadržava popis opreme s kojom je omogućena operacija zrakoplova (letenje) prema specifičnim uvjetima (okolišni uvjeti, ograničenje visine, brzine, itd.) ukoliko se ustanovi da je ta oprema neispravna prije polijetanja. MEL listu izrađuje zračni prijevoznik i ona se razlikuje za svaki tip zrakoplova te treba biti odobrena od Nacionalne agencije za civilno zrakoplovstvo.

procesu održavanja zrakoplova, ponekad nije moguće holistički sagledati sve podatke i donijeti pravovremenu odluku.

U ovoj doktorskoj disertaciji će se istražiti tehnike i alati kojima je moguće ukloniti prethodno navedene probleme kako bi se u procesu održavanja zrakoplova olakšalo odlučivanje i planiranje.

## **1.2. Motivacija i pregled dosadašnjih istraživanja**

Zrakoplovni sektor karakteriziraju visoki operativni troškovi, stroge regulacije i kompleksan proces planiranja [1]. U 2014. godini svjetska zrakoplovna flota za civilno zrakoplovstvo sastojala se od 24 597 zrakoplova, a zračni prijevoznici su potrošili 62,1 milijarde dolara na održavanje, popravak i remont (eng. *maintenance, repair and overhaul*, MRO) [2]. Predviđa se da će do 2018. godine svjetska zrakoplovna flota dodatno narasti te se očekuje da će broj zrakoplova u svjetskoj floti biti oko 28 000 [3].

Može se zaključiti da će s porastom broja zrakoplova i povećanjem potreba za zračnim prometom rasti i troškovi održavanja, dok će zračni prijevoznici kontinuirano tražiti nove načine za smanjenje troškova i povećanje raspoloživosti zrakoplova kako bi bili konkurentni na tržištu.

U posljednjih nekoliko desetljeća, strategije održavanja te njihov odabir doživjele su povećano zanimanje zbog sve veće važnosti upravljanja održavanjem. U 50-tim godinama prošlog stoljeća korektivno i reaktivno održavanje<sup>3</sup> (eng. *run-to-failure*) bile su dominantne strategije održavanja. U 60-tim godinama zahvati zamjene komponenata postali su planirani kako bi se izbjegli neplanirani zastoji bez obzira na „zdravstveno“ stanje komponenata te se iz takvog načina planiranja zahvata razvilo preventivno održavanje (eng. *time or use-based maintenance*). U drugoj polovici 80-tih godina tvrtke su počele obraćati pozornost na životni vijek komponenata što je pogodovalo razvoju dijagnostike po stanju, a time i strategije održavanja po stanju (eng. *Condition Based Maintenance*, CBM). Prema europskom standardu EN 13306:2010 održavanje po stanju se definira kao „preventivno održavanje koje podrazumijeva kombinaciju nadgledanja stanja i/ili inspekciju i/ili testiranje, analizu i odgovarajuće zahvate održavanja.“ Drugim riječima, aktivnosti održavanja za određenu komponentu odrađuju se na temelju prošlih informacija o stanju prateći njenu degradaciju ili neispravno ponašanje. U novije vrijeme,

---

<sup>3</sup> Reaktivno održavanje se provodi kada je kvar na komponenti/sustavu uzrokovao operacijski zastoj, a korektivno održavanje kada je uočena nepravilnost u radu komponente/sustava koja utječe na to da sustav ima sporiji odziv.

metode prognostike koje se bave predviđanjem pojava prije njihovog nastanka imaju sve veću primjenu u području održavanja tehničkih sustava. Na tim temeljima razvija se prediktivno održavanje (eng. *Predictive Maintenance*, PdM) koje se definira kao „održavanje po stanju koje se provodi na temelju prognoze proizašle iz ponavljajuće analize ili poznatih karakteristika i vrednovanja značajnih parametara vezanih uz degradaciju komponente“ (EN 13306:2010) [4].

Međunarodno udruženje zračnih prijevoznika (eng. *International Air Transport Association*, IATA) u svojoj godišnjoj analizi navodi da je primjenom novih tehnologija i razvojem trendova u održavanju u zrakoplovstvu (npr. praćenje pogrešaka i upozorenja preko sustava za nadzor tehničke ispravnosti zrakoplova (eng. (i dalje) *aircraft health monitoring system*, AHMS) i analiza velike količine podataka (eng. *big data*); korištenjem e-tehničke knjige zrakoplova (eng. *e-technical logbook*); primjenom dronova i/ili senzora za daljinske inspekcije, itd.), moguće smanjiti troškove održavanja, popravaka i remonta između 15 i 20%. IATA također procjenjuje da je prediktivnim održavanjem moguće povećati raspoloživost zrakoplova do 35% [2].

Sukladno tome, analizom nedavno objavljene znanstvene literature iz područja održavanja zrakoplova, utvrđeno je da su glavni optimizacijski ciljevi smanjenje troškova održavanja [1, 5-8], učinkovito korištenje materijalnih i/ili ljudskih resursa potrebnih za planiranje aktivnosti održavanja [8-10] te povećanje iskoristivosti, tj. raspoloživosti zrakoplova [11-14]. Glavni nedostatak navedenih optimizacijskih modela je to što i dalje nisu primjenjivi na sve zračne prijevoznike budući da koriste specifična ograničenja karakteristična za pojedine tipove zračnih prijevoznika (raznolikost i veličina zrakoplovne flote, ljudski i materijalni resursi, plan letenja, itd.).

Ono što karakterizira proces održavanja zrakoplova kod svih zračnih prijevoznika je da postoje velike količine podataka koji se svakodnevno bilježe u obliku pilotskih izvještaja, izvještaja o aktivnostima održavanja, zapisa o nezgodama i zastojsima, itd.

Znanja dobivena na temelju povijesnih podataka se mogu potencijalno koristiti za unapređenje procesa održavanja, no ekstrakcija tih znanja se ne može obaviti ručno. Tehnike dubinske analize podataka (eng. *data mining*) omogućavaju automatiziranu ili polu-automatiziranu ekstrakciju znanja iz baza podataka [15].

Zbog toga je, u novije vrijeme u predmetnoj literaturi uočen trend razvoja prediktivnih modela baziranih na tehnikama dubinske analize podataka.

R. S. Rodrigues i suradnici [16] navode kako je jedan od glavnih problema kod analize podataka o kvarovima u zrakoplovstvu nestrukturirani zapis tih kvarova. Primjenom tehnika dubinske analize teksta (eng. *text mining*) moguće je izlučiti (ekstrahirati) tekstualne uzorke (eng. *patterns*) povezane sa zapisima o kvarovima. Na temelju izlučenih tekstualnih uzoraka, uz primjenu umjetnih neuronskih mreža, spomenuti autori su izradili model za predviđanje raspoloživosti zrakoplova.

Gürbüz i suradnici [17] u svom istraživanju koriste zapise o zrakoplovnim nezgodama te primjenjuju stablo odlučivanja (eng. *decision tree*) za pronalazak odnosa i pravila između značajki odabranih teorijom približnih skupova (eng. *rough set theory*) i zapisa o ozbiljnim nezgodama (eng. *fatal incidents*).

Zapise o zrakoplovnim nezgodama koriste i Arockia i Appavu alias Balamuruga [18] te na njima primjenjuju pet različitih klasifikacijskih tehnika (naivni Bayes (eng. *Naive Bayes*, NB), stablo odlučivanja (eng. *decision trees*, DT), postupke temeljene na potpornim vektorima, algoritam *k*-najbližih susjeda i umjetne neuronske mreže) u kombinaciji sa statističkim tehnikama koje služe za pronalaženje odnosa između promatranih značajki. Autori zaključuju da je kombinacija analize glavnih komponentata (eng. *principal component analysis*, PCA) zajedno s klasifikacijskom tehnikom stablo odlučivanja najprikladnija za predviđanje zrakoplovnih nezgoda.

PCA analizu za redukciju značajki i zadržavanje samo bitnih značajki iz heterogenih operativnih podataka koriste i Bect i suradnici [19]. Svrstavanjem značajki u grupe sličnih objekata koristeći algoritam *k*-sredina (eng. *k-means*) te primjenom kanoničke diskriminativne analize (eng. *canonical discriminant analysis*, CDA), razvijeni su modeli za identifikaciju neispravnog funkcioniranja različitih sustava.

Come i suradnici [20] koriste samoorganizirajuće mape za predviđanje nenormalnih pojava (eng. *anomaly detection*) na motorima zrakoplova koristeći podatke koji se bilježe na motorima tijekom leta (operativne zapise). U ovome istraživanju samoorganizirajuće mape su poslužile kao pomoćni alat za vizualizaciju podataka zabilježenih tijekom leta na motoru s ciljem olakšanog dugoročnog nadgledanja i predviđanja kvarova, te za bolju organizaciju procesa održavanja motora.

Osim procjene pouzdanosti zrakoplovnih motora, García Nieto i suradnici [21] bave se predviđanjem preostalog životnog vijeka motora koristeći hibridni model koji kombinira

heuristički algoritam optimizaciju roja čestica (eng. *particle swarm optimization*, PSO) i algoritam temeljen na potpornim vektorima (eng. *support vector machines*, SVM). Autori zaključuju da hibridni PSO–SMV model u odnosu na SVM model, značajno poboljšava generalizacijsku sposobnost predviđanja kvarova.

Lv i suradnici [22] predviđaju trend trošenja motora (eng. *engine wear trend prediction*) temeljem spektralne analize ulja iz uzoraka prikupljenih prilikom svake izmjene ulja. Analizom koncentracije željeza (Fe element) u ulju, može se utvrditi signifikantna razina željeza što upućuje na određene probleme kod motora, kao što su istrošenost košuljice cilindra motora (eng. *cylinder liner*), problemi s pumpom za ulje (eng. *oil pump*), hrđa u sustavu, itd. Koristeći SVM metodu, autori zaključuju da je moguće previdjeti trošenje motora.

Wang i suradnici [23] predlažu novi pristup za analizu tehničke ispravnosti zrakoplovnih motora koji može poslužiti ujedno pilotima (on-line) i stručnjacima za održavanje (off-line) za lakšu analizu senzorskih podataka i dijagnozu stanja motora. Navedeni pristup koristi kombinaciju PCA i SVM metoda temeljem kojih je moguće selektirati relevantne senzorske podatke i zatim previdjeti nastanak kvara.

Za predikciju kvarova u stvarnom vremenu Chen i suradnici [24] u svom istraživanju opisuju razvoj softverske platforme koja povezuje sastavne module dubinske analize podataka (procesuiranje podataka, izlučivanje značajki, dijagnozu anomalija i prognozu zastoja). Koristeći Bayesovu teoriju za procjenu, generički okvir za filtriranje značajki integriran je unutar sustava za dijagnozu anomalija i predviđanje kvarova. Sustav je testiran na dva različita primjera, na ležajevima i istosmjernom motoru bez četkica, te se pokazalo da je njime moguće predvidjeti nastanak kvara.

Bayesove mreže koriste Ferreira i suradnici [25] te predlažu okvir za donošenje konačnih odluka vezanih uz poduzimanje zahvata održavanja na linijskom održavanju, tj. zamjene linijskih komponenata uzimajući u obzir preostali životni vijek komponente. Predloženi okvir koristi OSA-CBM (eng. *Open Systems Architecture for Condition-Based Maintenance*) standard koji definira faze (prikupljanje ulaznih podataka, manipulacija podacima, detekcija stanja, procjena stanja, prognoziranje, potpora pri odlučivanju i prezentacija rezultata) za obradu podataka u sustavu za nadgledanje stanja.

OSA-CBM standard za uspostavu okvira za predviđanje stanja tehničke ispravnosti sustava hidraulike koriste i Du i suradnici [26]. Kao ulaz koriste senzorske signale koji prolaze daljnju

obradu, tj. ekstrahiraju se relevantni senzorski podaci koji kasnije prolaze kroz faze detekcije, dijagnoze i prognoze. Autori navode kako je unutar tih faza moguće implementirati različite tehnike za prognoziranje, kao što su neuronske mreže, algoritam temeljen na potpunim vektorima, vremenske serije, Markov model, itd. Cilj ovog pristupa je da olakša planiranje aktivnosti u procesu održavanja.

L. R. Rodrigues i suradnici [27] analiziraju senzorske podatke primjenjujući heurističku metodu optimizacija kolonijom mrava (eng. *ant colony optimization*, ACO) s ciljem predviđanja preostalog životnog vijeka sustava (eng. *system-level remaining useful life*, S-RUL). Autori ističu kako nije nužno da kvar jedne komponente u sustavu dovodi do zastoja rada cijelog sustava. S ciljem minimiziranja troškova, u radu je predstavljen okvir za odabir kritičnih komponenta sustava koje trebaju biti pravovremeno zamijenjene kako bi sustav bio na željenoj razini sigurnosti. Ideja rada je da se odjednom zamijeni više komponentata na kojima je uočena nepravilnost u radu, kako ne bi došlo do zastoja u radu cijelog sustava. Navedeni pristup je primjenjiv na komponente koje su redundantne.

Viniacourt i suradnici [28] također ističu važnost analize senzorskih podataka za predviđanje neplaniranih zahvata održavanja. U radu koriste izvještaje o prošlim aktivnostima održavanja i senzorske podatke. Na senzorskim podacima primjenjuju klasifikacijske tehnike NB, SVM,  $k$ -NN i DT te predviđaju anomalije koje rezultiraju neplaniranim zahvatima održavanja. Rezultati klasifikacijskih modela su pokazali da DT u većini slučajeva najbolje predviđa pojavu nepravilnosti u radu sustava.

Desforges i suradnici [29] navode da je životni vijek zrakoplovnih sustava već poznat te je potrebno predvidjeti nastanak kvarova na razini komponentata. Samim time, autori su razvili model koji se temelji na objektno orijentiranim Bayesovim mrežama (eng. *Object Oriented Bayesian Networks*, OOBN), a kao ulaz koristi podatke o pouzdanosti komponentata. Svrha modela je da služi kao potpora pri odlučivanju u procesu održavanja zrakoplova s ciljem predviđanja kvarova na komponentama i planiranja prijevremenih zahvata održavanja.

Baptista i suradnici [30] analiziraju poruke upozorenja prikupljene iz sustava za nadzor tehničke ispravnosti zrakoplova primjenjujući sedam različitih tehnika dubinske analize podataka:  $k$ -najbližih susjeda (eng. *k-nearest neighbour*,  $k$ -NN), regresijska stabla (eng. *regression trees*, RT), SVM, linearni Bayesov model (eng. *bayesian generalized model*, BM), GBRT tehniku (eng. *gradient boosting with regression trees*), linearnu regresiju i neuronske mreže. Cilj ovog istraživanja je predviđanje preostalog vremena do zamjene komponente prilikom prve pojave

poruke upozorenja koja indicira nepravilnost u radu te komponente. Razvijeni klasifikacijski modeli su testirani na primjeru ventila za ispuštanje zraka na motoru (eng. *engine bleed valve*, EBV) te se pokazalo da dobru predikciju preostalog životnog vijeka ventila daju tehnike RT, BM i SVM.

Za razliku od prethodno spomenutih klasifikacijsko-predikcijskih pravila, Han i suradnici [31] koriste pravila pridruživanja (eng. *association rules*) za pronalaženje skrivenih uzoraka kvarova i uzoraka posljedičnih kvarova (kvarova koji se vežu jedni na druge) pod različitim okolnostima, tj. scenarijima (tip misije, lokacija – unutrašnjost ili obala, tip zrakoplova). Osim za predviđanje kvarova, ovaj pristup može unaprijediti iskoristivost zrakoplova predviđajući nadolazeću potrebu za doknadnim dijelovima.

Julka i suradnici [32] koriste D-SIMSPAIR alat za prognoziranje potrebnih zaliha i određivanje optimalne razine zaliha u zrakoplovnoj industriji. Kao ulaz koriste podatke o pouzdanosti komponenata na temelju kojih predviđaju vrijeme nastanka kvara. U navedenom istraživanju cilj je bio staviti fokus na kritične komponente koje mogu utjecati na pravovremenu otpremu zrakoplova te osigurati dostupnost tih komponenata na skladištu.

Pregledom objavljenih radova može se zaključiti da tehnike dubinske analize podataka još uvijek nisu dovoljno primijenjene u procesu održavanja zrakoplova, budući da zahtijevaju interdisciplinarna znanja kao što su razumijevanje baza podataka, statistička znanja te znanja o tehnikama strojnog učenja.

Nakon analize prikazanih istraživačkih pristupa iz literature uočeni su sljedeći nedostaci:

1. Za izradu klasifikacijskih modela i predviđanje određenog ishoda (kvar/preostali životni vijek komponente ili sustava) u procesu održavanja koristi se uglavnom jedan izvor podataka i nedostaje holistički pristup analize dostupnih izvora podataka.
2. Analiza poruka upozorenja prikupljenih iz AHMS-a uočena je samo u radu [30], no navedeni pristup, kao i u većini radova, je orijentiran na razinu komponente, a ne na razinu sustava.
3. Budući da su operativni zahtjevi zrakoplova u svakoj fazi leta različiti, sistemska logika kvarova se također razlikuje od faze do faze leta [33]. Samim time, potrebno je analizirati poruke upozorenja po fazama leta te promotriti njihov utjecaj na pojavu zapisa o nepravilnostima u radu sustava.



4. Prilikom selekcije značajki (eng. *feature selection*) s ciljem smanjenja ulaznog broja značajki, u predmetnoj literaturi se većinom primjenjuje PCA analiza. PCA analiza transformira podatke na takav način da definira manji broj novih značajki (komponenti) u čijem prostoru transformirani podaci zadržavaju većinu informacija iz izvornog prostora. To rezultira gubitkom tumačenja značajki koji je nužna posljedica transformacije podataka u niži prostor [34]. Stoga je, prilikom selekcije značajki, potrebno primijeniti druge postupke za smanjenje dimenzionalnosti podataka.
5. Većina klasifikacijskih algoritama u istraživanjima odabrana je bez dubljeg razumijevanja podataka, tj. nije provedena eksplorativna analiza podataka.
6. Potencijal analize nestrukturiranih tekstualnih zapisa primjenom tehnike dubinske analize teksta uočen je samo u radu [16].

Iz navedenog proizlazi motivacijski scenarij predloženog istraživanja: želi se doprinijeti daljnjem razvoju ovog područja izradom klasifikacijskih modela, temeljenih na tehnikama dubinske analize podataka, za potporu u procesu odlučivanja o raspoloživosti zrakoplova kako bi se u budućnosti bolje organizirao proces održavanja.

### **1.3. Cilj istraživanja, hipoteza rada i očekivani znanstveni doprinosi**

Cilj istraživanja je izraditi modele, temeljene na tehnikama dubinske analiza podataka, koji će služiti kao potpora pri odlučivanju o raspoloživosti zrakoplova, a pri modeliranju će se koristiti tri izvora podataka: podaci prikupljeni iz sustava za nadzor tehničke ispravnosti stanja zrakoplova, zapisi o nepravilnostima u radu sustava/prošlim kvarovima i zapisi o zastojsima u eksploataciji zrakoplova (operacijskim zastojsima).

Hipoteza istraživanja glasi: *Otkrivanjem značajki koje opisuju ponašanje kritičnih sustava zrakoplova u različitim fazama leta zrakoplova, na temelju tehnika dubinske analize podataka, moguće je predvidjeti raspoloživost zrakoplova.*

Očekuje se da će predloženo istraživanje rezultirati sljedećim znanstvenim doprinosima:

1. Formulacija postupka za strukturiranje podataka prikupljenih iz sustava za nadzor tehničke ispravnosti zrakoplova te postupka integracije tih podataka s drugim izvorima podataka.
2. Identificiranje i vrednovanje relevantnih značajki po svim fazama leta na razini kritičnih sustava zrakoplova.

3. Izrada novih klasifikacijskih modela za pouzdanije predviđanje ponašanja kritičnih sustava zrakoplova.

## 1.4. Terminologija područja

Prije nastavka istraživanja potrebno je definirati osnovne termine koji čine okosnicu ovog rada, kao što su raspoloživost zrakoplova i dubinska analiza podataka.

Prema NATO ARMP-7 (eng. *allied reliability and maintainability publication*) standardu [35], *raspoloživost* je definirana kao:

*„Sposobnost sustava da izvršava potrebnu funkciju u zadanim uvjetima rada u određenom vremenskom intervalu, pod pretpostavkom da su svi vanjski resursi dostupni.“*

Sukladno toj definiciji, Andela [36] navodi kako je raspoloživost zrakoplova indikator učinkovitosti koji se definira kao udio vremena u kojem zrakoplov može ispuniti svoju namjensku funkciju.

Dubinska analiza podataka definira se kao ekstrakcija skrivenih znanja i uzoraka iz setova podataka. Smatra se ključnim korakom procesa poznatog pod nazivom otkrivanje znanja u bazama podataka (eng. (i dalje) *Knowledge Discovery in Databases*, KDD), koji je definiran kao netrivialni proces za prepoznavanje novih i razumljivih uzoraka u velikim setovima podataka [37, 38].

U odjeljku 1.1 opisana je veza između nepravilnosti u radu sustava (eng. *fault*) i kvarova (eng. *failure, defect*). Definicije koje objašnjavaju ova dva termina u skladu su s definicijama objašnjenim od strane IFAC (eng. (i dalje) *International Federation of Automatic Control*, IFAC) tehničkog odbora koje su razvijene kako bi se postavio standard u ovome području i izbjegle nejasnoće među istraživačima [39].

Tehnički odbor IFAC-a definira ova dva termina na sljedeći način:

- *nepravilnosti u radu* – odstupanje barem jednog karakterističnog svojstva ili parametra sustava od prihvatljivog/uobičajenog stanja;
- *kvar ili defekt* – prekid rada sustava koji utječe na sposobnosti sustava da obavlja potrebnu funkciju u određenim radnim uvjetima.

Iz prethodno navedenih definicija može se razlučiti da je kvar stanje sustava koje je ozbiljnije od nepravilnosti u radu sustava. Primjerice, ukoliko se na aktuatorskoj uočila određena nepravilnost, on će još uvijek obavljati svoju funkciju iako će imati sporiji odziv i biti manje učinkovit [39].

Kao što je prethodno navedeno u odjeljku 1.1, nepravilnosti u radu zrakoplovnih sustava ili kvar sustava rezultirat će porukama upozorenja. Stoga je važno naglasiti razliku između poruka upozorenja i klasičnih senzorskih signala.

Iako su poruke upozorenja ekstrapolirane iz senzorskih signala kombinacijskom logikom, one se razlikuju od tih podataka. Prvo, senzorski signali su sinkroni, u smislu da se periodički aktiviraju. Suprotno tome, poruke upozorenja su događaji koji se generiraju asinkrono. Drugo, senzorski signali su kontinuirane značajke, dok su poruke upozorenja kategoričke. Treće, iako senzorski signali mogu imati nedostajuće i šumovite podatke (eng. *noisy data*), oni tvore skup podataka o trenutnom stanju sustava (eng. *ground-truth data*). Drugim riječima, ti podaci služe za izravno promatranje (monitoring) sustava i mogu se lako protumačiti. Za razliku od toga, poruke upozorenja podložne su daljnjem procesu interpretacije [30].

Naposljetku, iznesena je definicija o procesu odlučivanja (eng. *decision-making*) [40]:

*„Odlučivanje je mentalni proces u kojem osoba (ili grupa ljudi ili organizacija) uočava koja odluka treba biti donesena, sakuplja i procjenjuje informacije o mogućnostima te odabire najbolju mogućnost između više ponuđenih.“*

## **1.5. Struktura rada**

Disertacija je organizirana u šest poglavlja. U drugome poglavlju dan je pregled istraživanja povezanih s podjelom prognostičkih pristupa u procesu održavanja zrakoplova. Prikazani su koraci procesa za otkrivanje znanja u bazama podataka te su opisane različite tehnike nadziranog i nenadziranog učenja. Također su opisani postupci selekcije relevantnih značajki, mjere za evaluaciju klasifikacijskih modela, pristupi za optimizaciju modela i alati za dubinsku analizu podataka. U trećem poglavlju su opisani prikupljeni podaci te su razvijeni postupci za strukturiranje i udruživanje skupova podataka, nakon čega je provedena eksplorativna analiza podataka. Četvrto poglavlje prikazuje procese, razvijene unutar RapidMiner platforme, za otkrivanje relevantnih značajki i izgradnju klasifikacijskih modela prve i druge grupe. Rezultati klasifikacijskih modela prve i druge grupe su prikazani i evaluirani u petom poglavlju. Šesto poglavlje prikazuje zaključak disertacije i smjernice za daljnja istraživanja.

Disertacija sadržava i četiri priloga. U prilogu A je prikazan izvorni kôd algoritma za izdvajanje poruka upozorenja. Prilog B prikazuje izvorni kôd algoritma za prepoznavanje zapisa iz skupa podataka o prošlim kvarovima/nepravilnostima u radu sustava nastalih temeljem poruka upozorenja. Prilog C daje prikaz povezanih operatora unutar procesa razvijenih RapidMiner alatom za dubinsku analizu podataka. U prilogu D se nalazi rang značajki za analizirane kritične sustave.

## 2. PROGNOZIRANJE U PROCESU ODRŽAVANJA ZRAKOPLOVA

U ovome poglavlju opisana je podjela prognostičkih pristupa u procesu održavanja zrakoplova. Detaljno objašnjenje procesa otkrivanja znanja, s naglaskom na tehnike dubinske analize podataka prikazano je u odjeljku 2.3. Temeljem spoznaja iz odjeljka 2.3, u odjeljku 2.4 prikazane su daljnje smjernice za nastavak istraživanja u ovoj disertaciji.

### 2.1. Uvodna razmatranja

Zrakoplovi se sastoje od međusobno povezanih sustava čija funkcionalnost s napretkom tehnologije raste, a time se povećava i rizik od kvarova, što dovodi do poteškoća u održavanju. Kako bi se riješio taj problem, zračni prijevoznici trebaju uspostaviti arhitekturu za upravljanje i prognozu stanja tehničke ispravnosti zrakoplova (eng. (i dalje) *Prognostics and Health Management*, PHM) [41].

Glavni cilj za uspostavu PHM arhitekture nije izravno eliminacija kvarova, već predviđanje trenutka kada će se kvar dogoditi kako bi se poduzeli prijevremeni zahvati održavanja. Arhitektura za predviđanje stanja tehničke ispravnosti zrakoplova omogućava otkrivanje, dijagnozu, izolaciju parametara, prognozu i izvještavanje o nepravilnostima u radu te poduzimanje ispravnih zahvata održavanja na lociranoj komponenti u određenom trenutku vremena [26].

Hu i suradnici [42] navode kako dijagnoza tehničke ispravnosti zrakoplova i predviđanje kvarova ima važnu ulogu u procesu poboljšanja sigurnosti leta i smanjenja troškova održavanja zrakoplova.

Velika količina materijalnih i ljudskih resursa je uložena u razvoj tehnologije kako bi napredni sustavi za dijagnozu i predviđanje kvarova bili ugrađeni u nove generacije zrakoplova kao što su Boeing 787, Airbus 380 i neke nove generacije borbenih zrakoplova poput F-22 i F-35 [42, 43]. Međutim, ti sustavi u starijim generacijama zrakoplova (npr. Airbus 320, Boeing 737) zaostaju za novom tehnologijom te je zato potrebno poduzeti dodatne mjere za razvoj istraživanja u tom području.

## 2.2. Podjela prognostičkih pristupa

Najopsežniju podjelu prognostičkih pristupa izlažu Elattar i suradnici [44] klasificirajući prognostičke pristupe u četiri kategorije: pristupi zasnovani na pouzdanosti sustava (eng. *reliability based approach*); na fizičkom modelu sustava (eng. *physics-based approach*); zasnovani na podacima (eng. *data-driven approach*) i kombinirani pristupi (eng. *hybrid approach*).

Prva kategorija, tj. pristupi bazirani na pouzdanosti sustava većinom su vezani uz sustave i komponente koji nisu kritični. Ovaj pristup se temelji na analizi povijesnih zapisa o kvarovima pojedine komponente i na temelju toga se izračunava srednje vrijeme između dva kvara. Prednost ovog pristupa je što je jednostavan i služi za određivanje fiksnih intervala zamjene komponente prije nego što se očekuje da će doći do kvara. No, prijevremena zamjena komponenta je i nedostatak samog pristupa jer nisu razmotreni drugi utjecajni faktori, kao što su specifični operativni uvjeti komponente (eng. *component-specific conditions*) [44].

Fizički modeli sustava su matematički modeli koji opisuju modalitet kvarova i pojavu degradacije komponenta. Za formulaciju ovog modela potrebno je temeljito razumijevanje načina rada sustava/komponente. Nedostatak ovog pristupa je taj što je razvoj ovakvog modela za predviđanje preostalog životnog vijeka komponente skup, dugotrajan i računalno intenzivan proces koji ponekad ne može biti ostvaren. Zbog navedenih nedostataka, u novije se doba za prognoziranje sve više koriste pristupi bazirani na povijesnim podacima [44].

Pristup zasnovan na podacima smatra se ispravnom tehnikom prognoziranja za slučaj kada nije moguće koristiti fizičke modele zbog kompleksnosti sustava. Ideja ovog pristupa je da se koriste podaci prikupljeni tijekom operativnog vremena (npr. vibracije, temperaturne promjene) te da se ti podaci koriste za izradu modela koji će procjenjivati preostali životni vijek komponente. Ovi modeli su bazirani na tehnikama dubinske analize podataka. Prednost ovog pristupa je ta što nisu potrebna znanja o promatranom sustavu/komponenti, brzo se može implementirati, jednom razvijen pristup za određeni sustav se može primijeniti i na druge sustave te je moguće otkriti skrivena znanja u načinu rada sustava. Nedostaci ovog pristupa su ti što prikupljeni podaci mogu biti nedostadni ili šumoviti, interpretacija rezultata ponekad nije intuitivna jer nedostaju znanja o samom sustavu i većina tehnika dubinske analize podataka je aproksimativna [44].

Hibridni pristup je kombinacija pristupa baziranog na fizičkom modelu sustava i pristupa baziranog na podacima. Pristup se bazira na udruživanju podataka o fizikalnom stanju sustava/komponente i podataka prikupljenih tijekom operativnog vremena sustava/komponente i na taj način se pokušava predvidjeti preostali životni vijek sustava/komponente [44].

U slične kategorije, samo pod drugim nazivima, Hu [42] dijeli literaturu za dijagnozu i predviđanje kvarova na zrakoplovima temeljem informacija prikupljenih pomoću snimača podataka o letu (eng. (i dalje) *flight data recorder*, FDR). Navodi da se u tom području koriste uglavnom tri postupka: postupci temeljeni na analitičkoj redundanciji (eng. *analytical redundancy*), postupci koji predviđaju trendove parametara (eng. *parameter trend prediction*) i postupci temeljeni na tehnici stablo kvarova (eng. *fault tree*). Osnovna ideja primjene postupaka analitičke redundancije je da se prikupe senzorski podaci i izradi matematički model koji će opisivati različite mehanizme kvarova. Kada matematički model podsustava ili komponente može biti jednostavno postavljen, tada se preporuča korištenje postupaka analitičke redundancije. Za slučaj kada je mehanizam pojave kvara na podsustavu ili komponenti teško opisati, preporuča se korištenje postupaka koji predviđaju trendove parametara. Ova metoda se temelji na podacima, tj. promatraju se povijesni podaci prikupljeni iz FDR-a s trenutnim podacima iz FDR-a i predviđa se degradacija i preostali životni vijek komponenata. Za slučaj kada je logička veza unutar strukture podsustava jasna i vjerojatnost kvara svake komponente je poznata, koriste se pristupi temeljeni na tehnici stablo kvarova. Primjenom tehnike stabla kvarova moguće je izraditi logičku vezu između nepravilnosti u radu sustava i kvarova na komponentama.

Od prethodno navedenih kategorija iz istraživanja [44], Yousuf i suradnici [45] prognostičke pristupe vezane uz strukturne sustave zrakoplova dijele na dvije kategorije: bazirane na fizičkom modelu i bazirane na podacima. Istu podjelu prognostičkih pristupa vezanih uz predikciju kvarova ili predviđanje preostalog životnog vijeka sustava motora zrakoplova daju i Lv i suradnici [22].

U radu [45] se također navodi da su prognostički pristupi, koji se temelje na fizičkim modelima česti, no u nekim slučajevima teško ostvarivi zbog kompleksnosti systemske strukture. Stoga zaključuju da je predviđanje budućeg stanja tehničke ispravnosti pouzdanije primjenom metoda temeljenih na podacima.

Xiongzi i suradnici [46] i Pecht i Jaai [47] prognostičke pristupe vezane uz sustave zrakoplova dijele u tri kategorije: bazirane na fizičkom modelu, na podacima i hibridni pristup. Pecht i Jaai

[47] navode iste prednosti tih pristupa kao i prethodni istraživači, no preporučaju korištenje hibridnog pristupa jer taj pristup omogućava efektivno korištenje informacija kombiniranjem prednosti prva dva pristupa. Autori također naglašavaju kako razumijevanje fizičkog procesa i načina rada sustava može pomoći kod odabira tehnika za dubinsku analizu podataka.

## 2.3. Dubinska analiza podataka

Kao što je prikazano u odjeljku 1.2, literaturom u procesu održavanja u zrakoplovstvu obuhvaćeno je niz tehnika dubinske analize podataka kojima je zadatak otkrivanje asocijativnih pravila, grupa sličnih primjera (eng. *clustering*), klasifikacije primjeraka (eng. *classification*) i detekcije anomalija. Sve tehnike dubinske analize podataka mogu se svrstati u dvije kategorije učenja - u nadzirano i nenadzirano učenje, a cilj im je otkrivanje znanja u skupovima podataka.

### 2.3.1. Proces otkrivanja znanja u bazama podataka

Postupak dubinske analize podataka pruža širok spektar raspoloživih tehnika i alata podupirući time KDD proces koji se fokusira na razvoj matematičkog modela za određene primjene [48].

KDD proces se sastoji od nekoliko koraka, bez obzira na domenu i zadatak istraživanja [49]:

1. *Razumijevanje domene primjene.* Ovaj korak uključuje usvajanje znanja o problemu i otkrivanje ciljeva koji se žele postići KDD procesom.
2. *Sakupljanje (stvaranje) skupova podataka.* U ovome koraku se odabiru značajke i broj uzoraka u skupu podataka koji će poslužiti za otkrivanje znanja. Ovaj korak obično uključuje stvaranje upita (eng. *query*) u bazi podataka.
3. *Čišćenje i preprocesuiranje podataka.* Ekstremne, šumovite i nedostajuće vrijednosti u podacima je potrebno ukloniti kako ne bi utjecale na točnost modela. U nekim slučajevima je potrebno i balansirati skupove podataka.
4. *Redukcija i/ili transformacija podataka.* Ovaj korak se sastoji od pronalaženja relevantnih značajki primjenom metoda za smanjenje dimenzionalnosti i transformacije s ciljem pronalaženja nepromjenjivog (eng. *invariant*) prikaza podataka.
5. *Odabir zadatka procesa dubinske analize podataka.* Potrebno je definirati da li se treba provesti otkrivanje grupa sličnih primjera, detekcija anomalija, itd.
6. *Odabir algoritma za dubinsku analizu podataka.* Odabir klasifikacijskih algoritama (npr. naivni Bayes) ili algoritama za otkrivanje grupa sličnih primjera (npr. algoritam k-sredina).



7. *Dubinska analiza podataka*. Ovaj korak generira uzorke (eng. *patterns*) u određenom reprezentativnom obliku (npr. klasifikacijska pravila, stabla odlučivanja, regresijski modeli, itd.).
8. *Interpretacija rezultata*. Evaluacija dobivenih rezultata statističkim mjerama i prikaz rezultata u određenom obliku (vizualnom ili tabličnom).
9. *Konsolidiranje otkrivenog znanja*. Posljednji korak sastoji se od uključivanja otkrivenih znanja u sustav procjene ili odlučivanja te dokumentiranje i izvještavanje zainteresirane strane.

U KDD procesu je uobičajeno da se 80% vremena posveti odabiru podataka, preprocesuiranju i transformaciji podataka zbog složenosti heterogenih podataka, a preostalih 20% posveti dubinskoj analizi podataka, interpretaciji i evaluaciji rezultata [50].

Većina ovih koraka je iterativna i interaktivna i predstavlja uobičajene korake u KDD procesu. No, s gledišta provedbe rješenja, određeni KDD procesi mogu se provesti na mnogo različitih načina te je poznato da najbolji univerzalni pristup ne postoji [49].

### 2.3.2. *Nenadzirano učenje*

Nenadzirano učenje podrazumijeva tehnike dubinske analize podataka kojima je zadatak u većini slučajeva otkrivanje grupa sličnih primjera (provođenje klaster analize) ili pronalaženje asocijativnih pravila.

#### 2.3.2.1. *Klaster analiza*

Za razliku od tehnika nadziranog učenja, skup podataka koji će se koristiti u klaster analizi nema ciljne značajke. Iz tog razloga podaci će biti grupirani prema sličnim vrijednostima koje se nalaze unutar ulaznih značajki.

Drugim riječima, klaster analiza se može opisati kao proces otkrivanja sličnih uzoraka u skupu podataka i sređivanje tih uzoraka u grupe, tj. klastere. Otkriveni obrasci se uobičajeno prikazuju kao grupe točka u višedimenzionalnom prostoru [51].

Konačni rezultat klaster analize je deskriptivan budući da je svrha otkrivanje skupova sličnih uzoraka [52].

Primjer gdje se ova tehnika može koristiti u procesu održavanja zrakoplova je kod otkrivanja grupa zrakoplovnih sustava prema frekvencijama kvarova. Umjesto provođenja ručne analize

na podacima o prošlim kvarovima, moguće je provesti klaster analizu i tako otkriti grupe sustava (ili komponenata unutar sustava) koji se češće kvare.

Rasprostranjeni algoritmi grupiranja su hijerarhijsko grupiranje (npr. dendogram) i particijsko grupiranje (npr. algoritam  $k$ -sredina).

### 2.3.2.2. Asocijativna pravila

Ova se metoda sastoji u otkrivanju i prepoznavanju *asocijativnih pravila* ili *pravila pridruživanja* koja ukazuju na učestalu vezu između vrijednosti značajke i pojedine značajke u zadanom skupu podataka [53]. Često se koristi u analizi potrošačke košarice (eng. *market basket analysis*) u transakcijskim sustavima, gdje se pokušava odgovoriti na pitanja oblika: *što će kupiti korisnik X, ako je kupio proizvode A i B?*

Općenito, asocijativno pravilo se zapisuje u obliku:

$$A \Rightarrow B \text{ ili } A_1 \wedge \dots \wedge A_m \rightarrow B_1 \wedge \dots \wedge B_n, A \cap B = \emptyset,$$

gdje su  $A_i$  i  $B_j$  parovi *značajka-vrijednost značajke*.

S obzirom da neki transakcijski sustav koji prati ponašanje kupca može pronaći na tisuće, pa i milijune takvih pravila, definiraju se mjere koje će pokazati koliko je neko pravilo zanimljivo analitičaru. Takve mjere su podrška i pouzdanost pravila [53]:

- (i) *podrška pravila* (eng. *support*):  $\text{podrška}(A \Rightarrow B) = P(A \cup B)$ ,  
Ako je podrška pravila  $p\%$ , to znači da u skupu transakcija  $p\%$  njih sadrži i  $A$  i  $B$ .
- (ii) *pouzdanost pravila* (eng. *confidence*):  $\text{pouzdanost}(A \Rightarrow B) = P(B | A)$ ,  
Ako je pouzdanost pravila  $r\%$ , to znači da u skupu transakcija  $r\%$  onih transakcija koje sadrže  $A$ , sadrže i  $B$ .

Općenito, samo ona pravila koja ispunjavaju minimalno zadani prag podrške i prag pouzdanosti smatraju se zanimljivima. Takva se pravila nazivaju *jaka pravila*.

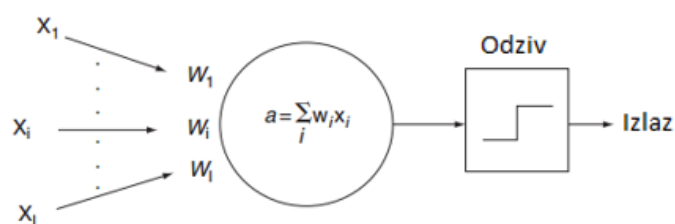
### 2.3.3. Nadzirano učenje

Nadzirano učenje je tehnika dubinske analize podataka koja koristi označene podatke (eng. *labeled data*) koji predstavljaju klase (razrede) svakog uzorka. Temeljem spoznaja prikupljenih iz odjeljka 1.2., najčešće korištene tehnike, tj. klasifikacijski algoritmi, koji spadaju u ovu kategoriju su stablo odlučivanja (DT), naivni Bayes (NB), metoda potpornih vektora (SVM) i neuronske mreže (NN).

#### 2.3.3.1. Neuronske mreže

Neuronske mreže korištene za kvantificiranje temelje se na ranom razumijevanju strukture i funkcije ljudskog mozga. Drugim riječima, neuronske stanice primaju električne impulse iz susjednih stanica i akumuliraju ih sve dok se ne prekorači granična vrijednost (prag). Nakon toga, stanice predaju impuls susjednoj ćeliji. Kapacitet ćelije za pohranjivanje električnih impulsa i prag se kontroliraju biokemijskim procesima koji se mijenjaju tokom vremena. Ova promjena je pod kontrolom autonomnog živčanog sustava i primarno je sredstvo pomoću kojeg čovjek uči razmišljati ili aktivirati vlastito tijelo [54].

Umjetni neuroni u mrežama (slika 2-1) ugrađuju ova dva procesa i mijenjaju ih numerički, a ne biokemijski. Proces sakupljanja (eng. *aggregation process*) obično prihvaća ulazne podatke na način da ih zbraja. Proces aktivacije je predstavljen određenom matematičkom funkcijom, obično linearnom ili logističkom. Linearne aktivacijske funkcije najbolje funkcioniraju za numeričke procjene (tj. regresiju), dok logistička aktivacijska funkcija najbolje opisuje problem klasifikacije [54].

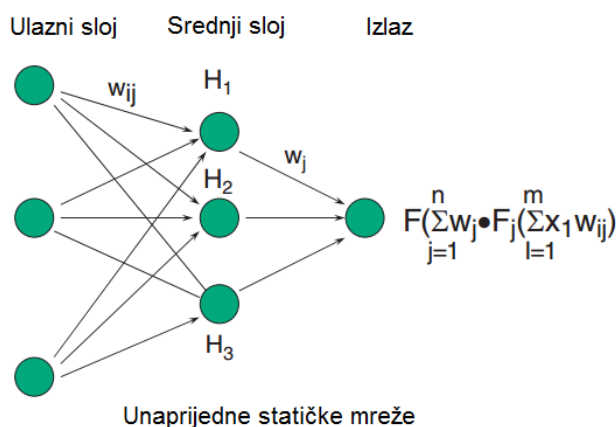


Slika 2-1. Arhitektura neurona s  $X_i$  ulaznih značajki [54]

Oznaka  $X_i$  označava ulazne značajke, predstavljajući broj neurona povezanih s ljudskim neuronima. Oznaka  $W_i$  predstavlja težine povezane sa svakom vezom, koje su analogne jačini povezanosti. Jačina veze predstavlja blizinu između dva neurona u dijelu koji se naziva sinapsa [55].

Umjetni neuroni povezani su zajedno u arhitekturu koja tvori mrežu u kojoj je svaka ulazna varijabla  $X_i$  (zvana ulazni čvor (eng. *input node*) povezana s jednim ili više izlaznih čvorova. Ova mreža se zove umjetna neuronska mreže. Kada je ulazni čvor povezan s izlaznim čvorom preko funkcije zbroja i logističke aktivacijske funkcije, matematička obrada je analogna logističkoj regresiji s kategoričkim izlazom. Takva konfiguracija neuronske mreže predstavlja snažnu klasifikacijsku tehniku koja ima sposobnost prepoznavanja nelinearnih odnosa između izlaznih i ulaznih značajki.

Zanimljive mogućnosti neuronske mreže dolaze do vidjela kada se u nju uvrsti i srednji sloj neurona (eng. *middle layer*), kao što je prikazano na slici 2-2.



Slika 2-2. *Arhitektura neuronskih mreža* [54]

Težine ( $W_{ij}$ ) se dodjeljuju svakoj vezi između ulaznih i srednjih čvorova, kao i vezama između srednjih i izlaznih čvorova. Ove težine imaju sposobnost modeliranja nelinearnih odnosa između ulaznih čvorova i izlaznih čvorova, u čemu i leži vrijednost neuronske mreže za rješavanje problema dubinske analize podataka. Čvorovi u srednjem sloju omogućavaju modeliranje nelinearnih odnosa između ulaznih čvorova i izlaznih čvorova (odluke).

Što je veći broj čvorova u srednjem sloju, veća je i sposobnost neuronske mreže da prepozna nelinearne obrasce u skupu podataka. Međutim, s povećanjem broja čvorova u srednjem sloju, eksponencijalno se povećava i vrijeme učenja modela (eng. *training time*), što može rezultirati prenućenošću (eng. *overtraining*) modela. Takav model može dobro predviđati na podacima za učenje, no loše će predviđati na drugom skupu podataka. Nažalost, ne postoje empirijska pravila za definiranje broja čvorova u srednjem sloju. Jedina smjernica je da se koristi više čvorova za slučaj kada postoji dovoljan broj uzoraka i velik broj ulaznih značajki u skupu podataka za

učenje modela. Drugim riječima, ukoliko je dani klasifikacijski problem kompleksan, preporuča se veći broj čvorova u srednjem sloju [54].

Neuronska mreža može ujedno predvidjeti kontinuirane značajke i diskretne značajke (da/ne ili 1/0). Alternativno, mrežna arhitektura može biti takva da sadržava više izlaznih čvorova za procjenu ili klasifikaciju te se čak može upotrijebiti i kao algoritam za grupiranje podataka (eng. *clustering algorithm*).

Karakteristike NN tehnike su [54]:

- Mogu riješiti probleme s velikim brojem značajki i klasificirati objekte čak i kada je distribucija uzoraka u  $N$ - dimenzionalnom parametarskom prostoru vrlo složena.
- Dobro prepoznaju nelinearne odnose između značajki.
- Distribucija podataka nije bitna za izgradnju modela.
- Mogu biti spore, posebice kod izrade modela, ali i u fazi primjene već izgrađenog modela.
- Teško je shvatiti logiku kako mreža „odlučuje“ te je teško fizikalno objašnjenje modela (eng. *black box*).

#### **2.3.3.2.    *Stablo odlučivanja***

Tehnika stablo odlučivanja predstavlja hijerarhijsku grupu odnosa između značajki organiziranih u strukturu drva, počevši od jedne značajke (poput debla) koja se naziva izvorni čvor (eng. *root node*). Taj čvor se dalje dijeli u dvije ili više grana, predstavljajući odvojene klase (razrede) korijena čvora (za slučaj kada je značajka kategorička) ili specifične raspone unutar čvorova (za slučaj kada je značajka kontinuirana) [54].

Prilikom svakog grananja čvorova, postavlja se određeno „pitanje“ čiji odgovor rezultira određenom raspodjelom klase ili raspona značajke. Primjer pitanja može biti „Da li ovaj parametar utječe na zastoj komponente?“ Takva, i slična pitanja se koriste za izgradnju stabla odlučivanja s kategoričkim podjelama.

Stabla odlučivanja također mogu biti izgrađena s više podjela. Pitanja koja se postavljaju na svakom čvoru za grananje definiraju se u smislu neke mjere nečistoće (eng. *impurity measure*), odražavajući ujednačenost nastalih slučajeva u podjeli.

Svaka grana je dalje podijeljena s obzirom na klase ili raspone drugih značajki. Kod svake podjele, čvor koji se dijeli naziva se *čvor roditelj* (eng. *parent node*), a čvorovi koji nastaju iz

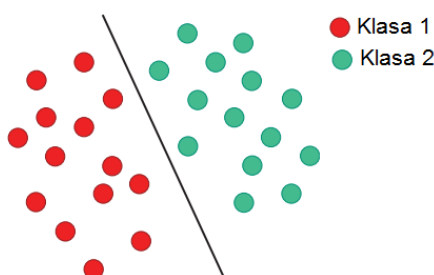
tog čvora se nazivaju *čvorovi djeca* (eng. *child node*). Ovaj se postupak nastavlja dok se ne zadovolji određeno pravilo za zaustavljanje, kao što je minimalni broj slučajeva koji se treba nalaziti u konačnom čvoru (eng. *terminal leaf node*). Navedeni proces se naziva rekurzivna podjela (eng. *recursive partitioning*).

Karakteristike DT tehnike su [54]:

- Stabla odlučivanja posebno su prikladna za kategoričku klasifikaciju.
- Tehnika je korisna za estimaciju kompleksnih problema zbog jednostavnosti prikaza i dobre tolerancije na ekstremne vrijednosti (eng. *outliers*).
- Zahtijeva veliki broj uzoraka za učenje i izgradnju modela.
- Kod podjele značajki koje su kontinuirane, teško se donosi granica podjele.

### 2.3.3.3. *Metoda potpornih vektora*

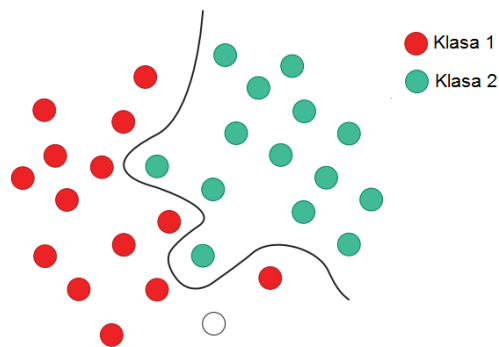
Metoda potpornih vektora se bazira na konceptu statističke teorije učenja koji definira ravnine kao granice odlučivanja. Ravnina odlučivanja dijeli uzorke koji imaju različite klase, kao što je prikazano na slici 2-3 [54].



Slika 2-3. *Prikaz linearnog razdvajanja prostora ulaznih podataka* [54]

Prilikom toga, linija odvajanja zapravo definira granicu između uzoraka koji se svrstavaju u različite klase – na lijevoj strani su uzorci koji spadaju u klasu crveno (klasu 1), dok su na desnoj strani uzorci koji spadaju u klasu zeleno (klasu 2).

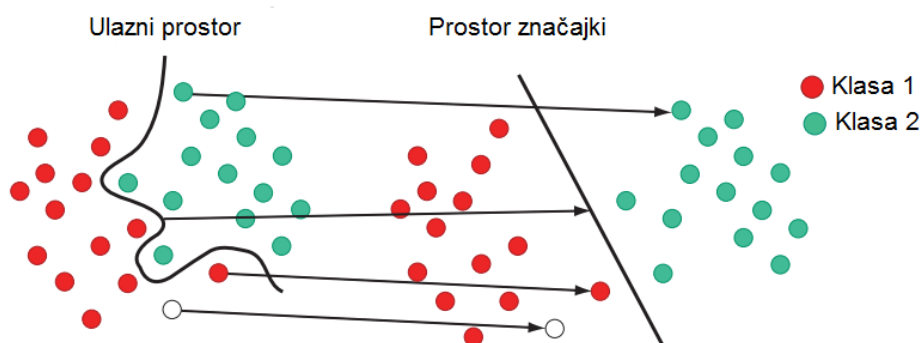
Međutim, većina klasifikacijskih problema nije toliko jednostavna i često je složenija struktura potrebna za optimalno izdvajanje, tj. pravilno razvrstavanje novih uzoraka (verifikacijski podataka) na temelju dostupnih primjera (podataka za učenje modela). Na slici 2-4 je prikazano da potpuna separacija crvenih i zelenih uzoraka zahtijeva granicu koja je složenija od linije, tj. zapravo je određena krivulja.



Slika 2-4. Prikaz nelinearnog razdvajanja prostora ulaznih podataka [54]

Klasifikacijski zadaci koji se temelje na određivanju linija (granica) između uzoraka različitih klasa poznatiji su kao klasifikacijski algoritmi hiper-ravnina (eng. *hyperplane classifiers*) [54]. U navedene algoritme spada i SVM metoda.

Slika 2-5 prikazuje osnovnu ideju SVM metode.



Slika 2-5. Prikaz svrstavanja ulaznih uzoraka u prostor značajki gdje su razredi linearno razdjeljivi [54]

Na lijevoj strani slike prikazani su izvorni uzorci koji su svrstani u dvije klase korištenjem skupa matematičkih funkcija poznatih kao jezgre (eng. *kernels*). Proces svrstavanja uzoraka je poznat kao mapiranje (eng. *mapping*) ili transformacija uzoraka u novi prostor (hiper-prostor, eng. *hyperspace*) različitih dimenzija koji se naziva prostor značajki (eng. *feature space*). U novom prostoru, svrstani uzorci su linearno odvojivi i stoga, umjesto izgradnje složene krivulje razdvajanja, potrebno je pronaći optimalnu liniju razdvajanja koja može odvojiti uzorke u dvije klase (crvenu i zelenu). Također, hiper-prostor se može konfigurirati tako da maksimizira marginu oko hiper-ravnine koja razdvaja klase podataka [54].

Za izgradnju optimalne hiper-ravnine, SVM koristi algoritam iterativnog učenja (eng. *iterative training algorithm*), minimizirajući time grešku (eng. *error function*).

SVM metoda podržava ujedno regresijske i klasifikacijske probleme, kao i kontinuirane i diskretne značajke.

#### 2.3.3.4. *Naive Bayes*

Kod većine klasifikacijskih problema postoji vrlo malo znanja o novom entitetu koji je potrebno razvrstati u određenu klasu (razred). Stoga je prihvatljivo da se klasifikacijski proces sagleda kao subjektivna odluka (eng. *judgmental call*). Slijedeći ovu logiku, NB algoritam je postao jedna od ključnih tehnika dubinske analize podataka.

NB algoritam se bazira na Bayesovom teoremu koji je opisan na sljedeći način [53]:

Neka je svaki uzorak iz skupa  $D$  podataka predstavljen kao  $n$ -dimenzionalni vektor:  $X = (x_1, \dots, x_n)$ , gdje  $x_i$  vrijednost značajke  $A_i$  za zadani uzorak. Neka je zadano  $m$  klasa:  $C_1, \dots, C_m$ . Broj uzoraka u skupu  $D$  je  $d$ , a u pojedinoj klasi  $d_i$ . Ako za uzorak  $X$  nije poznata klasa, tada će klasifikator predvidjeti da  $X$  pripada onoj klasi koja ima najveću aposteriornu vjerojatnost  $P(C_i | X)$ , tj.  $X$  će biti razvrstan u klasu  $C_i$  ako i samo ako vrijedi:  $P(C_i | X) > P(C_j | X)$  uz uvjet  $j = 1, \dots, m$  i  $i \neq j$ . Klasa  $C_i$  za koju je  $P(C_i | X)$  maksimalna zove se *maksimalna aposteriori hipoteza*. Prema Bayesovom teoremu vrijedi:

$$P(C_i | X) = \frac{P(X | C_i)P(C_i)}{P(C_i)}. \quad (2.1)$$

S obzirom da je  $P(X)$  konstantan za sve razrede, potrebno je maksimizirati  $P(X | C_i)P(C_i)$ .

Ako unaprijed nisu poznate vjerojatnosti pojedinih klasa, onda je uobičajeno pretpostaviti da je vjerojatnost pripadanja pojedinoj klasi jednaka, tj. da vrijedi:  $P(C_1) = \dots = P(C_m)$ , pa je stoga potrebno maksimizirati  $P(X | C_i)$ , a ako ne vrijedi da je vjerojatnost pripadanja pojedinoj klasi jednaka, onda je potrebno maksimizirati jednakost  $P(X | C_i)P(C_i)$ .

Pripadnost pojedinoj klasi može se odrediti kao:

$$P(C_i) = \frac{d_i}{d}. \quad (2.2)$$

Uz *naivnu* pretpostavku o neovisnosti značajki, određuje se  $P(X | C_i)$  kao:

$$P(X | C_i) = \prod_{k=1}^n P(x_k | C_i), \quad (2.3)$$

a vjerojatnosti  $P(x_k | C_i)$  mogu se odrediti na temelju podataka iz skupa pokaznih uzoraka:

- (i) Ako je značajka  $A_k$  kategorička značajka, onda je:



$$P(x_k | C_i) = \frac{d_{ik}}{d_i}, \quad (2.4)$$

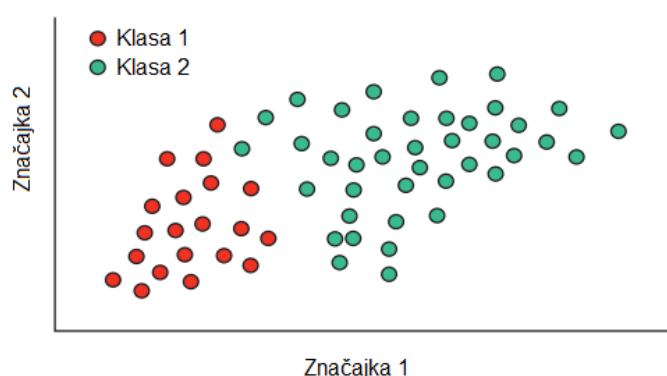
gdje je  $d_{ik}$  broj uzoraka iz pokaznog skupa uzoraka kojima je  $x_k$  vrijednost značajke  $A_k$ , a pripadaju razredu  $C_i$ .

(ii) Ako je  $A_k$  značajka s *kontinuiranim* vrijednostima, onda se pretpostavlja da značajka ima Gaussovu distribuciju, tj. da vrijedi:

$$P(x_k | C_i) = \frac{1}{\sqrt{2\pi}\sigma_{C_i}} e^{-\frac{(x_k - \mu_{C_i})^2}{2\sigma_{C_i}^2}}, \quad (2.5)$$

gdje je  $\mu_{C_i}$  srednja vrijednost i  $\sigma_{C_i}$  standardna devijacija značajke  $A_k$  za razred  $C_i$ .

Kako bi se objasnio koncept NB algoritma, na slici 2-6 je prikazan skup podataka u kojemu su uzorci klasificirani unutar klasa prema svojim značajkama.



Slika 2-6. Prikaz uzoraka podijeljenih u dvije klase [54]

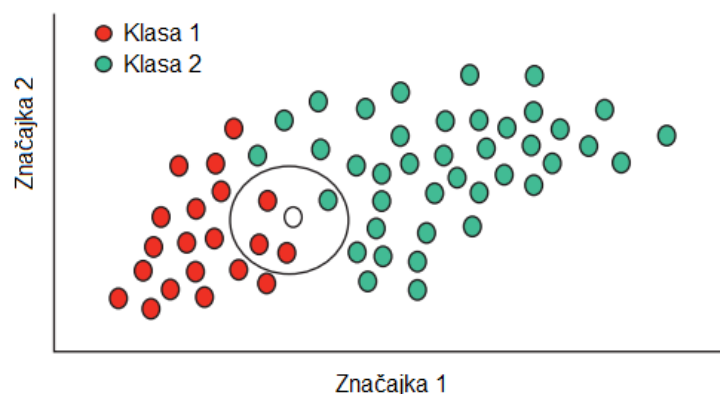
S obzirom na dosadašnju podjelu uzoraka na slici 2-6 u dvije klase (crvenu - klasa 1 i zelenu – klasa 2), cilj je da se novi uzorci svrstaju u točne klase. Na temelju činjenice da postoji dvostruko više uzoraka koji pripadaju klasi 2, može se pretpostaviti da je vjerojatnost dva puta veća kako će novi uzorak biti klasificiran u klasu 2 umjesto u klasu 1. Prema NB teoremu, ovo uvjerenje je poznato kao vjerojatnost prognoze (eng. *prior probability*). Vjerojatnost prognoze se temelji na „dokazima“ iz prošlog klasifikacijskog postupka, što u ovome slučaju predstavlja postotak uzoraka u klasi 1 i klasi 2, i može se koristiti za predviđanje klasifikacijskog ishoda novih uzoraka.

Budući da sveukupno u klasi 1 i klasi 2 postoji 60 uzoraka, od čega 40 pripada klasi 2 (zelena klasa), a 20 klasi 1 (crvena klasa), vjerojatnosti prognoze za svaku klasu mogu se prikazati kao [54]:

*Vjerojatnost prognoze da će uzorak pripasti klasi 2 je proporcionalna omjeru (40/60)*

*Vjerojatnost prognoze da će uzorak pripasti klasi 1 je proporcionalna omjeru (20/60)*

Na temelju formulacije vjerojatnosti prognoze, potrebno je svrstati novi uzorak  $X$  (bijeli krug) u ispravnu klasu, kako je prikazano slikom 2-7.



Slika 2-7. Pozicija novog uzorka za koji je potrebno odrediti klasu [54]

Budući da su prošli uzorci ispravno klasificirani, može se pretpostaviti da što je više uzoraka iz klase 2 (ili klase 1) u blizini uzorka  $X$ , to je veća vjerojatnost da će novi uzorak pripasti upravo toj klasi. Kako bi se procijenila ta vjerojatnost (eng. *likelihood*), potrebno je razmotriti područje oko uzorka  $X$  (prikazano većim krugom na slici 2-7.), koje obuhvaća određeni broj (odabran *a priori*) prethodno klasificiranih uzoraka neovisno o njihovoj klasi. Na temelju toga moguće je izračunati vjerojatnost kako slijedi [54]:

*Vjerojatnost da će  $X$  pripasti klasi 2 proporcionalna je omjeru (1/40)*

*Vjerojatnost da će  $X$  pripasti klasi 1 proporcionalna je omjeru (3/20)*

Iako se iz prikaza vjerojatnosti procjene može zaključiti kako bi  $X$  trebao biti svrstan u klasu 2 (s obzirom da postoji dvostruko više uzoraka unutar te klase), temeljem prethodno navedenog izraza vidljivo je kako je veća vjerojatnost da  $X$  bude svrstan u klasu 1.

U NB algoritmu, konačna klasifikacijska odluka se donosi kombinacijom ovih dviju vjerojatnosti kako bi se izračunala aposteriorna ili uvjetna vjerojatnost (eng. *conditional probability*):

*Uvjetna vjerojatnost da će  $X$  pripasti klasi 2 jednaka je izrazu:  $(4/6) \times (1/40) = 1/60$*

*Uvjetna vjerojatnost da će  $X$  pripasti klasi 1 jednaka je izrazu:  $(4/6) \times (2/20) = 1/10$*

Finalno, uzorak  $X$  će biti svrstan u klasu 1 (crvenu) jer je izračunom dobivena veća uvjetna vjerojatnost za tu klasu.

Osim prethodno opisanog NB algoritma, Bermúdez [56] također navodi kako su istraživači predstavili različite modifikacije ovog algoritma; NB model s procjenom gustoće jezgre za podatke koji nemaju normalnu distribuciju [57] ili hijerarhijski NB model koji pretpostavlja da su značajke međusobno korelirane [58].

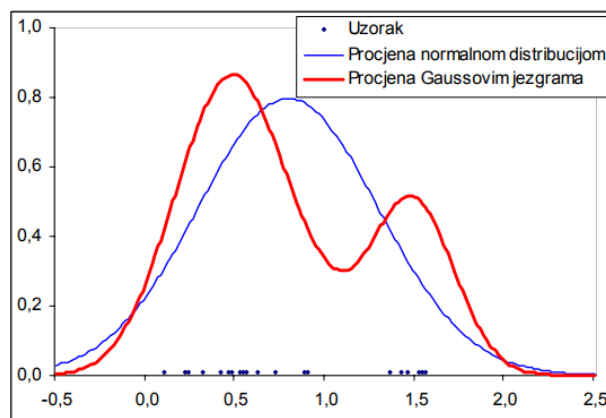
Modificirani NB model s procjenom gustoće jezgre je primjenjivan u literaturi iz razloga što se vrijednosti numeričkih značajki ne ponašaju uvijek po normalnoj razdiobi i tada se funkcija gustoće vjerojatnosti može procijeniti koristeći sumu više Gaussovih jezgri [34]. Procjena gustoće prikazana je izrazom [34, 57]:

$$P(x_j^{(k)} = x_{jk}, C_i) = \frac{1}{d} \sum_d G(x_{jk}, x_d, \sigma_{Ci}), \quad (2.6)$$

pri čemu je  $d$  broj primjeraka značajke  $x_j^{(k)}$  koja ima vrijednost ciljne značajke jednaku  $C_i$  na skupu za učenje, a  $x_d$  su konkretne vrijednosti značajke  $x_j^{(k)}$  na skupu za učenje.

Važan element metode ocjene gustoće vjerojatnosti korištenjem jezgri je odabir parametra širine  $\sigma$ . Može se pokazati da ova metoda ocjene gustoće vjerojatnosti pokazuje neka dobra teoretska svojstva ako se parametar  $\sigma$  približava nuli kako brojnost uzorka raste ka beskonačnosti [57, 59].

Slika 2-8 prikazuje gustoću vjerojatnosti procijenjenu normalnom distribucijom (tanka plava crta) u usporedbi s procjenom gustoće korištenjem Gaussovih jezgri (deblja crvena crta).



Slika 2-8. Usporedba metoda procjene gustoće vjerojatnosti numeričkih značajki [59]

Eksperimentalno je pokazano da je ova modifikacija NB algoritma vrlo korisna te u većini slučajeva gdje je narušena pretpostavka normalnosti performanse su znatno bolje, dok je utjecaj neznatan u slučajevima kad pretpostavka stoji [59].

Karakteristike NB tehnike su [54]:

- Usprkos svojoj jednostavnosti, NB algoritam može nadmašiti sofisticiranije klasifikacijske algoritme u točnosti predviđanja.
- Funkcija jezgre može biti modelirana na nekoliko različitih načina uključujući normalnu, lognormalnu, gamma i Poissonovu funkciju gustoće ili se može koristiti „estimacija jezgre“ (ne pretpostavlja jednu definiranu razdiobu).
- Dobro predviđa i na malim i velikim skupovima podataka za učenje.
- Pretpostavlja da ulazne značajke nisu međusobno korelirane, što nije čest slučaj u praksi.

#### 2.3.4. Postupci selekcije značajki

Performanse klasifikacijskih algoritama uvelike ovise o relevantnosti odabranih ulaznih značajki u odnosu na ciljanu značajku, koja se još naziva i klasa. U posljednjih nekoliko godina, veliki istraživački interes pojavio se u području poboljšanja performansi algoritama kroz smanjenje dimenzionalnosti podataka [60].

Iz tog razloga, u predmetnoj literaturi predložene su različite metode za selekciju značajki. Svrha tih metoda je odbacivanje nevažnih ili redundantnih značajki u određenom vektoru značajki.

Metode odabira značajki mogu se svrstati u dvije kategorije, filterske metode (eng. *filter methods*) i metode omotača (eng. *wrapper methods*) [61]. Kod filterskih metoda određuje se optimalni podskup značajki korištenjem općih karakteristika podataka. Drugim riječima, podskup značajki se odabire jednostavnim rangiranjem vrijednosti svake značajke prema funkciji evaluacije [62]. Za razliku od metoda omotača, filterske metode ne ovise o klasifikacijskim algoritmima i mogu se kombinirati s više različitih algoritama.

Metode omotača prvo koriste optimizacijski algoritam koji dodaje ili uklanja značajke da bi dobio različite podskupove značajki, a nakon toga koriste klasifikacijski algoritam koji procjenjuje taj podskup značajki [63]. Iako daju bolje rezultate u praktičnim primjenama, metode omotača su računalno zahtjevnije za obradu i sporije od filterskih metoda [61, 62 i 64].

Također, njihov nedostatak je taj što ne skaliraju dobro na velikim, visoko dimenzionalnim skupovima podataka [63].

Sukladno tim zaključcima, u ovome radu će se koristiti filterske metode za otkrivanje relevantnih značajki. Često primjenjivani postupci u literaturi, koji spadaju u filterske metode, su postupci gini indeks (eng. (i dalje) *gini index*, GI), informacijski dobitak (eng. (i dalje) *information gain*, IG), omjer informacijskog dobitka (eng. (i dalje) *information gain ratio*, IGR) i korelacija (eng. (i dalje) *correlation*, KP).

#### 2.3.4.1. *Informacijski dobitak*

Informacijski dobitak je postupak selekcije značajki koji se temelji se na radu Claudea Shannona u području teorije informacija gdje je proučavan „informacijski sadržaj“ poruke. Značajka s najvećom vrijednosti informacijskog dobitka minimizira potrebnu informaciju za klasifikaciju skupa. Takav pristup minimizira očekivani broj testova u procesu klasifikacije. Očekivana informacija potrebna za klasifikaciju u nekom skupu  $D$  je [63]:

$$Info(D) = -\sum_{i=1}^m p_i \log_2(p_i), \quad (2.7)$$

pri čemu je  $p_i$  vjerojatnost da proizvoljni element u  $D$  pripada klasi  $C_i$ , a procjenjuje se pomoću  $|C_{i,D}| / |D|$ . Koristi se log funkcija s bazom 2, jer su informacije kodirane binarno.  $Info(D)$  je samo prosječna potrebna količina informacija za identifikaciju oznake klase neke  $n$ -torke u skupu  $D$ . U ovom trenutku, sve informacije koje postoje temelje se isključivo na odnosu proporcije  $n$ -torki u svakoj klasi.  $Info(D)$  je također poznat kao entropija od  $D$ . Nadalje, vodeći se pretpostavkom da će se  $n$ -torke particionirati u  $D$  temeljem neke značajke  $A$  koja ima  $v$  različitih vrijednosti,  $\{a_1, a_2, \dots, a_v\}$  postavlja se pitanje: „Koliko informacija će još uvijek biti potrebno nakon particioniranja kako bi se došlo do točne klasifikacije?“ To se može izračunati sljedećim izrazom [63]:

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j). \quad (2.8)$$

Izraz  $|D_j| / |D|$  djeluje kao ponder  $j$ -te particije.  $Info_A(D)$  je očekivana količina informacija potrebna za klasifikaciju  $n$ -torke iz  $D$  nakon particioniranja po  $A$ . Manje potrebnih informacija znači veću čistoću particija. Informacijska dobit se definira kao razlika između izvornog zahtjeva za informacijama (tj. onog koji se temelji samo na udjelu klasa) i novog zahtjeva (tj. dobivenog nakon podjele po  $A$ ) [63]:

$$Gain(A) = Info(D) - Info_A(D). \quad (2.9)$$

$Gain(A)$  nam govori koliko bi se dobilo podjelom po  $A$ . Očekivano je smanjenje informacijskih potreba uzrokovano poznavanjem vrijednosti od  $A$ . Značajka  $A$  s višim informacijskim dobitkom, ( $Gain(A)$ ), je bolje rangirana po danim trening  $n$ -torkama.

#### 2.3.4.2. Omjer informacijskog dobitka

Postupak informacijskog dobitka preferira značajke s većim brojem vrijednosti. Kao primjer, može se sagledati značajka ID, koja je jedinstvena za svaki element u bazi. Podjela po toj značajki dovest će do onolikog broja podskupova koliko imamo primjera u skupu primjera za učenje, a svaki od podskupova imat će samo jedan element. Kako svaki element u takvim podskupovima pripada samo jednoj klasi, informacija koja će se dobiti jednaka je nuli, tj.  $InfoID(D) = 0$ . Dobit je u slučaju podjele po toj značajki maksimalna, i ta značajka dobiva najviši rang, no takva podjela nije korisna. Novi postupak koji pokušava prevladati pristranost iz navedenog slučaja naziva se omjer dobiti (engl. *gain ratio*). To se postiže uz svojevrsnu normalizaciju informacijske dobiti pomoću novodefinirane vrijednosti "*SplitInfo*" čija vrijednost je definirana analogno s  $Info(D)$  [63]:

$$SplitInfo_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} x \log_2 \frac{|D_j|}{|D|}. \quad (2.10)$$

Ova vrijednost uzima u obzir broj  $n$ -torki koje imaju određenu vrijednost značajke u odnosu na ukupan broj  $n$ -torki u  $D$ . Omjer dobiti se definira kao [63]:

$$Gain(A) = \frac{Gain(A)}{SplitInfo(A)}. \quad (2.11)$$

Analizirajući početni primjer s ID-ovima, može se zaključiti da će u tom slučaju  $GainRatio$  biti mali. Ukoliko postoji  $n$  primjera za učenje, koji spadaju u  $m$  klasa, uz uvjet da postoji puno više primjera za učenje nego klasa,  $GainRatio$  će biti malen, jer će informacijska dobit iznositi maksimalno  $\log_2(m)$  dok će  $SplitInfo$  imati vrijednost  $\log_2(n)$  koja je puno veća od  $\log_2(m)$ . Značajka s višim omjerom je bolje rangirana po danim trening  $n$ -torkama [63].

#### 2.3.4.3. Gini indeks

Ginijev indeks mjeri nečistoću podataka, particije podataka ili skupa trening  $n$ -torki  $D$  na sljedeći način [63]:

$$Gini(D) = 1 - \sum_{j=1}^v p_i^2, \quad (2.12)$$

gdje je  $p_i$  je vjerojatnost da  $n$ -torka iz  $D$  pripada klasi  $C_i$ , a procjenjuje se pomoću  $|C_{i,D}| / |D|$ . Zbroj se izračunava nad  $m$  klasa. Ginijev indeks za kategoričku značajku  $A$  koji dijeli  $D$  na particije  $D_1$  i  $D_2$  je [63]:

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2). \quad (2.13)$$

Smanjenje nečistoće nastalo po značajki  $A$  je:

$$\Delta Gini(A) = Gini(D) - Gini_A(D) \quad (2.14)$$

Značajka koja maksimizira smanjenje nečistoće (ili, ekvivalentno, ima minimalni Ginijev indeks) je najbolje rangirana po zadanim trening  $n$ -torkama. Ginijev indeks je postupak odabira značajki koji se temelji na stupnju nečistoće u rezultirajućim particijama. Za utvrđivanje ranga pojedine značajke potrebno je usporediti stupanj nečistoće skupa (prije podjele) sa stupnjem nečistoće rezultirajućih particija (nakon podjele). Pošto je nečistoća skupa prije podjele jednaka za sve značajke, maksimiziranje dobiti ekvivalentno je minimiziranju srednje vrijednosti mjera nečistoća rezultirajućih particija. Što je njihova razlika veća to je odabrana značajka bolja [63].

#### 2.3.4.4. Korelacija

Koeficijent korelacije je statistički test koji mjeri linearnu povezanost između dvije značajke. Vrijednost koeficijenta može biti između -1 i 1. Apsolutna vrijednost koeficijenta daje jačinu povezanosti i što je bliža 1, to znači da je povezanosti jača. Predznak povezanosti upućuje na smjer povezanosti. Pozitivni predznak upućuje na to da su značajke izravno povezane, dok negativni upućuje na obrnutu povezanost značajki [65].

U procesu otkrivanja relevantnih značajki, koeficijent korelacije služi za evaluaciju koliko točno ulazna značajka predviđa ciljnu značajku ili klasu, neovisno o ostalim ulaznim značajkama. Sve ulazne značajke su rangirane na temelju korelacijskog rezultata. Ako je poznata kovarijanca  $cov(x_i, y)$  između ulazne značajke  $x_i$  i ciljne značajke  $y$ , kao i njihove varijance  $var(x_i)$  i  $var(y)$ , korelacija se može izračunati preko izraza:

$$\rho = \frac{cov(x_i - y)}{\sqrt{var(x_i) \cdot var(y)}}. \quad (2.15)$$

Ova jednadžba može se koristiti samo u slučaju kada su vrijednosti kovarijanca i varijance poznate. Ukoliko su te vrijednosti nepoznate, estimacija korelacije se može dobiti na temelju Pearsonovog koeficijenta korelacije  $\rho$  na sljedeći način:

$$\rho = \frac{\sum_{k=1}^m (x_i^k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^m (x_i^k - \bar{x})^2 \sum_{k=1}^m (y_k - \bar{y})^2}}. \quad (2.16)$$

Formula koristi aritmetičku sredinu ulazne značajke i izlazne značajke za izračun koeficijenta korelacije za  $m$  uzoraka.

### 2.3.5. Evaluacija i optimizacija klasifikacijskih modela

Procjena performansi klasifikacijskih algoritama važna je ne samo za procjenu točnosti predviđanja, nego i za odabir najboljeg klasifikacijskog algoritma. Prilikom izgradnje klasifikacijskih modela važno je koristiti tehnike probira podataka kako bi se dobila procjena stvarne greške klasifikacijskog modela.

#### 2.3.5.1. Tehnike probira za vrednovanje performansi modela

Wong [66] navodi da su dva česta pristupa za vrednovanje performansi klasifikacijskih modela na skupu podataka  $k$ -struka unakrsna validacija i metoda izostavljanja (eng. *leave-one-out*).

Princip rada  $k$ -struke unakrsne validacije je da raspodijeli primjere za učenje u  $k$  odvojenih skupova  $D_i$ , gdje vrijedi da je  $i=1, \dots, k$ . Nakon toga, za svaki  $i$  od 1 do  $k$  koristi skup  $D_i$  kao skup podataka za testiranje, a ostale podatke u skupovima  $D_j$ ,  $j \neq i$  koristi za učenje modela. Na testnom skupu podataka  $D_i$  u svakoj iteraciji izračuna grešku modela i kao rezultat daje prosječnu grešku svih  $k$  modela.

Podjela primjera za učenje može biti slučajna (eng. *shuffled sampling*) i stratificirana (eng. *stratified sampling*). Za razliku od slučajnog probira, stratificirani način probiranja osigurava da su u svakom  $k$  skupu bude otprilike podjednak udio primjeraka iz obje klase.

Za razliku od  $k$ -struke unakrsne validacije, metoda izostavljanja iz skupa primjera za učenje  $D_i$  privremeno izdvoji primjer na kojemu kasnije testira model i daje prosječnu grešku modela. Taj primjer ne služi za učenje modela, već služi samo za verifikaciju.

Hu [67] navodi kako modeli bazirani na podacima većinom koriste metodu izdvajanja (eng. *holdout method*) koja dijeli skup podataka na dva međusobno isključiva podskupa – podskup za učenje modela i podskup za testiranje (izdvojeni podskup). Prednost metode izdvajanja je ta što je jednostavna i nije računalno zahtjevnja. Međutim, nedostatak je taj što generira veliku varijancu izlazne procjene i zahtijeva dodatan skup podataka za testiranje. Kako bi se prevladali



navedeni nedostaci, predlaže se korištenje  $k$ -struke unakrsne validacije za procjenu točnosti izgrađenog klasifikacijskog modela.

Na temelju prošlih istraživanja, Jović [34] navodi sljedeća empirijska pravila:

1. Za skupove veće od tisuću primjeraka po razredu koristi se metoda izdvajanja.
2. Za skupove između sto i tisuću primjeraka obično se koristi postupak unakrsne validacije s  $k$  preklopa.
3. Za male skupove koristi se uzorkovanje s ponavljanjem (engl. (i dalje) *bootstrap*) i metoda izostavljanja ili postupak unakrsne validacije s dva preklopa i pet ponavljanja.

### 2.3.5.2. Mjere za vrednovanje performansi modela

U postupku dubinske analize podataka, klasifikacijski algoritmi za predviđanje kategoričke klase (razreda) se u osnovi ocjenjuju matricom zabune koja se sastoji od četiri kategorije, kao što je to prikazano u tablici 2-1.

Tablica 2-1. Prikaz matrice zabune za kategoričku klasifikaciju [68]

		Stvarne vrijednosti	
		0	1
Predviđene vrijednosti	0	TN	FN
	1	FP	TP

Akronimi TN, FN, FP i TP predstavljaju sljedeće vrijednosti:

TN = (eng. *true negative*) je broj negativnih primjeraka koji su ispravno klasificirani kao negativni,

FN = (eng. *false negative*) je broj pozitivnih primjeraka koji su pogrešno klasificirani kao negativni,

FP = (eng. *false positive*) je broj negativnih primjeraka koji su pogrešno klasificirani kao pozitivni,

TP = (eng. *true positive*) je broj pozitivnih primjeraka koji su ispravno klasificirani kao pozitivni.

Iz matrice zabune proizlazi tablica 2-2 koja predstavlja fundamentalne mjere za vrednovanje performansi klasifikacijskog modela.

Tablica 2-2. Prikaz evaluacijskih mjera temeljenih na analizi matrice zabune [68]

Mjera (eng. <i>performance measure</i> )	Formula	Interpretacija
<i>Točnost</i>	$\frac{TP + TN}{TP + TN + FP + FN}$	Ukupna točnost klasifikacijskog modela
<i>Klasifikacijska pogreška =</i> <i>1-Točnost</i>	$\frac{FP + FN}{TP + TN + FP + FN}$	Ukupna pogreška klasifikacijskog modela
<i>Osjetljivost ili odziv</i>	$\frac{TP}{TP + FN}$	Točnost pozitivnih primjeraka
<i>Specifičnost</i>	$\frac{TN}{TN + FP}$	Točnost negativnih primjeraka
<i>Preciznost</i>	$\frac{TP}{TP + FP}$	Ispravnost klasificiranja pozitivnih primjeraka

*Točnost*, najčešća metrika za ocjenjivanje klasifikacijskih algoritama, procjenjuje ukupnu učinkovitost algoritma procjenom vjerojatnosti stvarne vrijednosti klase (pozivne ili negativne). Naspram toga, *klasifikacijska pogreška* je procjena vjerojatnosti pogrešne klasifikacije prema predviđanju modela.

*Preciznost* je mjera ispravnosti (eng. *correctness*), tj. od ukupnog broja pozitivnih primjeraka, koliko je stvarno pozitivnih primjeraka.

*Osjetljivost (odziv)* je mjera cjelovitosti ili točnosti pozitivnih primjera, tj. koliko primjeraka iz pozitivne klase (klasa 1) je ispravno označeno.

*Preciznost* i *osjetljivost* su mjere koje imaju inverzan odnos. No, za razliku od mjera *točnosti* i *klasifikacijske pogreške*, ove dvije mjere nisu osjetljive na promjene u distribuciji podataka [68].

*Specifičnost* je uvjetna vjerojatnost ispravno negativnih primjeraka (klasa 0) s obzirom na drugu klasu. Ova mjera aproksimira vjerojatnost da je negativna klasa pozitivna. Drugim riječima, mjere *osjetljivost* i *specifičnost* procjenjuju učinkovitost algoritma na jednoj klasi (pozitivnoj ili negativnoj).

U slučaju klasno neuravnoteženih skupova podataka (eng. *imbalanced data*), mjera *točnosti* stavlja veću težinu na većinsku klasu. To rezultira time da klasifikacijski model pogrešno

klasificira primjerke manje zastupljene klase koji su u nekim slučajevima od veće važnosti za domenski problem [68].

Iz tog razloga, razvijena je dodatna mjera, *F-mjera*, koja kombinira *osjetljivost (odziv)* i *preciznost*:

$$F = \frac{2 \cdot \text{osjetljivost} \cdot \text{preciznost}}{\text{osjetljivost} + \text{preciznost}} \quad (2.17)$$

*F-mjera* je definirana kao harmonijska sredina *preciznosti* i *osjetljivosti*. Vrijednost *F-mjere* raste proporcionalno s povećanjem *osjetljivosti* i *preciznosti*. Veća vrijednost *F-mjere* ukazuje na to da model dobro predviđa primjerke pozitivne klase [68].

Osim *F-mjere*, za procjenu točnosti obiju klasa, tj. *osjetljivosti* (točnosti pozitivnih primjeraka) i *specifičnosti* (točnosti negativnih primjeraka), razvijena je mjera *geometrijska sredina* (eng. *G-mean*):

$$G = \sqrt{\text{osjetljivost} \cdot \text{specifičnost}} \quad (2.18)$$

Ova metrika upućuje na balans između klasifikacijskih performansi manjinske i većinske klase. Loše performanse u predviđaju pozitivne klase će rezultirati malom vrijednošću *G-mjere*, čak i ukoliko su negativni uzorci ispravno klasificirani [68].

Opsežno istraživanje klasno neuravnoteženih skupova podataka prikazano je u radu [69] gdje se navodi da su *osjetljivost (odziv)*, *specifičnost*, *preciznost*, *F-mjera* i *geometrijska sredina* mjere koje dobro vrednuju problem neuravnoteženosti skupova.

### 2.3.5.3. Optimizacija modela

Klasifikacijski algoritmi se mogu dodatno podijeliti na parametarske i neparametarske probabilističke algoritme. Parametarski klasifikacijski algoritmi se temelje na statističkoj razdiobi vjerojatnosti (eng. *statistical probability distribution*) svakog razreda (klase), dok se neparametarski algoritmi koriste u slučaju nepoznate funkcije gustoće vjerojatnosti [70].

Određeni skup parametara, koje posjeduju klasifikacijski algoritmi, je potrebno optimizirati kako bi algoritam bolje pretraživao prostor rješenja.

Različiti pristupi, kao što su rešetkasto pretraživanje (eng. *grid search*) ili genetički algoritmi (eng. *genetic algorithms*), su prepoznati kao učinkoviti za pronalaženje odgovarajućih vrijednosti parametara za dani skup podataka [71].

Rešetkasto pretraživanje je standardni pristup optimizacije parametara koji je široko rasprostranjen [72]. Korištenjem ovog pristupa moguće je odrediti raspon i korake svakog parametra, nakon čega će svaka kombinacija parametara biti testirana kako bi se pronašla najbolja vrijednost parametra. Ukoliko postoji više parametara, prvo se pretraže vrijednosti iz jednog skupa parametara s fiksnim vrijednostima drugog parametra. Nakon što se odredi najbolja vrijednost tog parametra, ta vrijednost se fiksira i prelazi se na ostale parametre.

Genetički algoritmi predstavljaju pristup koji se temelji na biološkoj evoluciji jedinke. Populacija jedinke razvija se tijekom više generacija. Jedinke koje su prikladnije u odnosu na ostale jedinke u populaciji, one opstaju i križanjem i/ili mutacijom prenose svoj kôd na sljedeću generaciju koja se stvara, dok se stara uništava. Ovaj postupak se ponavlja dok se ne ostvari određeni uvjet za zaustavljanje evolucijskog procesa. U kontekstu optimizacije parametara, rješenje je jedan specifični skup svih parametara određenog klasifikacijskog algoritma. Prikladnost jedinke je opisana performansama (eng. *performance*) koje je ostvario klasifikacijski algoritam koristeći odabrane vrijednosti parametara [71].

Reif i suradnici [71] napominju kako je prednost rešetkastog pretraživanja ta što pronalazi točnija rješenja, no u slučaju velikog broja parametara dolazi do kombinatorne eksplozije. Za razliku od toga, genetički algoritmi mogu pronaći dobra rješenja za manje vremena, ali točnost ovog pristupa je obično niža od rješenja dobivenih rešetkastim pristupom.

#### 2.3.6. Alati za dubinsku analizu podataka

U području dubinske analize podataka postoje skupovi podataka s preko milijun uzoraka koje je potrebno analizirati. Stoga postoji potreba za automatiziranim alatima koji olakšavaju transformaciju i analizu tih uzoraka. Samim time, u novije vrijeme razvijeni su mnogi besplatni (eng. *open-source*) alati za dubinsku analizu podataka, kao što su RapidMiner (RM), Waikato Environment for Knowledge Analysis (WEKA), KNIME, RStudio, Orange, Tanagra, NLTK, itd. [73].

Chen i suradnici [74] proveli su anketu o besplatnim alatima za dubinsku analizu podataka kako bi usporedili prednosti i nedostatke tih alata. Temeljem provedenog istraživanja, autori zaključuju kako većina dostupnih besplatnih alata ima dobru funkcionalnost i pružaju niz različitih tehnika za istraživače iz raznih područja.

Detaljnije istraživanje o besplatnim alatima, koji ne zahtijevaju velike hardverske i softverske investicije, provode Al-Odan i Al-Daraiseh [75] te uspoređuju pet alata: RStudio, RapidMiner,

WEKA, KNIME i Orange. U istraživanju je fokus bio usporediti alate prema različitim kriterijima korisničkom evaluacijom (anketirani su korisnici iz IT područja) i tehničkom evaluacijom. Što se tiče tehničke evaluacije (usporedba sučelja, procesa instalacije, povezivanje s različitim izvorima podataka, usporedba dostupnih algoritama, itd.), može se zaključiti da su razlike među alatima vrlo male i zanemarive. No, prilikom korisničke evaluacije prema određenim kriterijima (intuitivnost, dostupnost edukacijskih materijala, dosljednost, itd.) pokazalo se da su RapidMiner i KNIME dva najbolje ocijenjena alata.

Kao prilog tome ide i istraživanje Gartner kuće koja je u svom izvještaju iz 2017. godine, pod nazivom „*Magic Quadrant for Data Science Platforms*“ [76], svrstala RapidMiner i KNIME alate u kvadrant lidera.

## **2.4. Zaključak poglavlja i smjernice za nastavak istraživanja**

Temeljem provedenih istraživanja u odjeljcima ovog poglavlja, zaključuje se sljedeće:

1. Zbog velikih količina podataka generiranih u procesu održavanja zrakoplova, modeli bazirani na podacima rezultiraju bržim otkrivanjem znanja. Navedeni modeli omogućavaju i veću istraživačku slobodu u smislu primjene različitih tehnika dubinske analize podataka.
2. Ključni koraci u KDD procesu su odabir podataka, preprocesuiranje i transformacija podataka. Drugim riječima, ulazni podaci su ključ uspješnog KDD procesa. Pogrešno odabrani i pripremljeni podaci neće rezultirati otkrivanjem korisnih znanja. Stoga je potrebno detaljno razraditi postupke strukturiranja i udruživanja skupova podataka.
3. Razumijevanje skupova podataka je važno kod odabira klasifikacijskih tehnika i algoritama. Budući da su karakteristike klasifikacijskih algoritama prikazane u odjeljku 2.3.3, u nastavku istraživanja je potrebno analizirati skupove podataka i odabrati klasifikacijski algoritam koji bi najbolje odgovarao domenskom problemu.
4. Filterski postupci selekcije značajki se mogu koristiti neovisno o klasifikacijskom algoritmu i ti postupci će se primijeniti u daljnjem istraživanju s ciljem izgradnje klasifikacijskog algoritma s boljim performansama.
5. Analogno zaključku iz točke 3., prije odabira tehnika probira i statističkih mjera za evaluaciju modela, potrebno je provesti eksplorativnu analizu podataka.

6. U slučaju odabira probabilističkog parametarskog klasifikacijskog algoritma, provest će se optimizacija parametara korištenjem rešetkastog pretraživanja jer je istraživanjem otkriveno da ovaj pristup pronalazi točnija rješenja.
7. Sukladno zaključcima iz odjeljka 2.3.6, zbog svoje funkcionalnosti i rasprostranjenosti, u disertaciji će se primarno koristiti RapidMiner platforma za dubinsku analizu podataka, što je u skladu s preporukom Gartner istraživanja [76]. Ulaz u procese razvijene unutar RapidMiner platforme, koji će predstavljati klasifikacijske modele, će biti skupovi podataka sastavljeni od podataka prikupljenih iz sustava za nadzor tehničke ispravnosti stanja zrakoplova, zapisa o nepravilnostima u radu sustava/prošlim kvarovima i zapisa o zastoјima u eksploataciji zrakoplova (operacijskim zastoјima).

### 3. OBRADA I ANALIZA PRIKUPLJENIH PODATAKA

U ovome poglavlju opisani su prikupljeni podaci na temelju koji će biti izgrađeni klasifikacijski modeli. S ciljem ostvarivanja znanstvenog doprinosa #1, u trećem odjeljku je razvijen postupak za strukturiranje i udruživanje skupova podataka. Udruživanjem skupova podataka, u četvrtom odjeljku je provedena eksplorativna analiza na temelju koje su dane smjernice za daljnje faze istraživanja.

#### 3.1. Uvodna razmatranja

Prediktivno održavanje često zahtijeva analizu podataka iz različitih izvora, što rezultira udruživanjem podataka (eng. *data fusion*). Udruživanje podataka je tehnika koja uključuje proces kombiniranja podataka iz različitih izvora s ciljem otkrivanja kvalitetnijih informacija nego što bi to bilo moguće analizom jednog izvora podataka [77].

Problem udruživanja podataka (eng. *data fusion*) predstavlja učestali istraživački interes koji se proteže i izvan domene civilnih zrakoplova. Brojni istraživači su predložili postupke za analizu uzročnih odnosa između podataka o trenutnom stanju komponenata (npr. vibracijskih signala) i pojave kvarova na ležajevima, lopaticama vjetrenjača ili prijenosnim mehanizmima (eng. *gearbox*) [78-82]. Dodatna istraživanja također predlažu postupke za povezivanje senzorskih podataka i strukturnih oštećenja na rotacijskim sustavima [83] ili na strojevima za drobljenje (eng. *mining mills*) [84].

Analogno tim istraživanjima, dio istraživanja u procesu održavanja zrakoplova opisuje upotrebu senzorskih podataka za predviđanje kvarova na kritičnim komponentama [85, 86].

Budući da zrakoplovni sustavi generiraju veliku količinu senzorskih podataka, razvijen je niz postupaka koji povezuju te podatke s podacima koji se bilježe tijekom aktivnosti održavanja. Sukladno tome, Bernardo [85] je dao pregled postupaka za udruživanje podataka koji služe za dijagnozu kvarova na temelju senzorskih podataka koje generiraju zrakoplovni sustavi. Isti autor predlaže pristup razvijen od strane JDL-a (eng. *Joint Directors of Laboratories*) koji se bavi udruživanjem i kombinacijom podataka iz većeg broja senzora koji su ugrađeni u zrakoplovima. JDL postupak predlažu i Hall i suradnici [87] za analizu senzorskih podataka te zaključuju kako je udruživanje važan korak prilikom faze preprocesuiranja jer velike količine nepovezanih podataka mogu utjecati na sposobnost rasuđivanja u procesu donošenja odluke.

Postupci i okviri za dijagnozu kvarova u zrakoplovu na temelju senzorskih podataka se također spominju u radovima [88-93].

Međutim, navedeni radovi ne spominju udruživanje poruka upozorenja, koje predstavljaju kategoričke značajke, s podacima iz drugih izvora, kao što su podaci o kvarovima i operacijskim zastoјima.

## **3.2. Opis dostupnih podataka**

Za doktorsku disertaciju prikupljeni su podaci u razdoblju od 76 uzastopnih mjeseci (15.08.2008.-31.12.2015.) za četiri zrakoplova tipa Airbus A319/20 koji pripadaju floti jednog zračnog prijevoznika.

Dostupni podaci za analizu obuhvaćaju tri nezavisna izvora podataka:

1. podatke prikupljene iz AHMS-a koji sadrže zapise o porukama upozorenja nastale na razini sustava po različitim fazama leta zrakoplova,
2. podatke o prošlim kvarovima ili nepravilnostima u radu sustava proizašle iz pilotskih zapisa, a koji su smješteni unutar informacijskog sustava za planiranje aktivnosti i resursa,
3. podatke o zastoјima u eksploataciji zrakoplova prikupljene iz mjesečnih analiza o pouzdanosti zrakoplovne flote.

Sukladno ATA 100 referentnom standardu [94] u istraživanju su korišteni podaci vezani uz sustave zrakoplova. Temeljem iskustvenih spoznaja, dobivenih od strane zračnog prijevoznika, provest će se analiza za pet kritičnih sustava zrakoplova; sustava automatskog leta (eng. (i dalje) *auto flight*, ATA 22), sustava upravljanja letjelicom (eng. (i dalje) *flight controls*, ATA 27), hidrauličkog sustava (eng. (i dalje) *hydraulic power*, ATA 29), sustava podvozja (eng. (i dalje) *landing gear*, ATA 32) i navigacijskog sustava (eng. (i dalje) *navigation*, ATA 34). Navedeni sustavi se smatraju „kritičnima“ jer je analizom mjesečnih izvještaja o pouzdanosti zrakoplovne flote uočeno da su, od svih sustava, navedeni sustavi uzrokovali najviše operacijskih zastoја na floti tog zračnog prijevoznika.

Iako su tri zrakoplova tipa A319, dok je jedan je tipa A320, unutar sva četiri zrakoplova su ugrađeni isti sustavi. Zrakoplovi su približno jednake starosti, tj. dva zrakoplova su u upotrebi od 1998. godine, jedan od 1999. godine i jedan od 2000. godine (tablica 3-1).



Tablica 3-1. *Popis dijela tehničkih karakteristika zrakoplova*

Zrakoplov*	Tip zrakoplova	Godina proizvodnje	Ukupan broj sati naleta**
Ac <sub>1</sub>	A319	1998	43 695
Ac <sub>2</sub>	A319	1998	43 409
Ac <sub>3</sub>	A320	1999	39 908
Ac <sub>4</sub>	A319	2000	38 445

\* Zbog zaštite podataka u radu će se umjesto izvornog registracijskog broja koristiti pseudo naziv za svaki zrakoplov

\*\* Ukupan broj sati naleta zrakoplova registriran na kraju razdoblja u kojemu su podaci prikupljeni (31.12.2015.)

Iz tablice 3-1 je vidljivo da je ukupan broj sati naleta približno jednak na svim zrakoplovima. Također je poznato da su svi zrakoplovi eksploatirani u gotovo identičnim uvjetima, tj. da lete na iste destinacije koje zračni prijevoznik omogućava, da su upravljani i održavani od strane istog osoblja te da imaju jednaki program održavanja budući da su isti sustavi ugrađeni unutar tih zrakoplova.

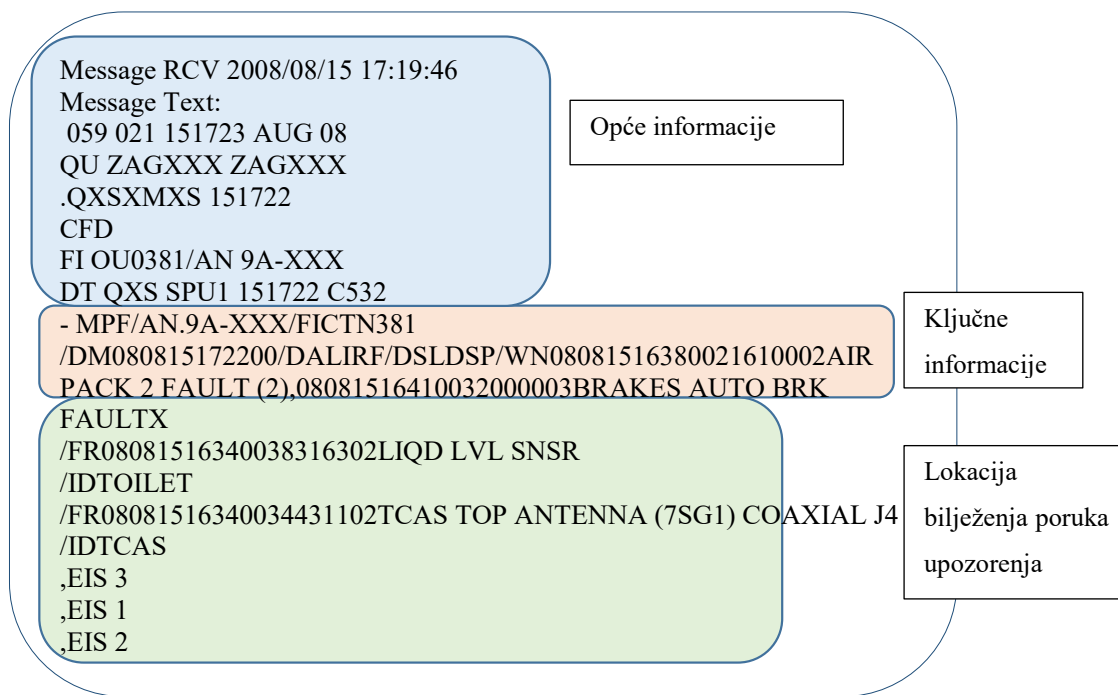
Iz svega navedenog može se zaključiti da se podaci prikupljeni sa sva četiri zrakoplova mogu sastaviti u jedan skup podataka, ovisno o izvoru podataka.

Tako će u daljnjoj analizi ulazni skup podataka poruka upozorenja prikupljenih iz AHMS-a (nadalje u tekstu: skup WN zapisa), skup podataka o prošlim kvarovima/nepravilnostima u radu sustava (nadalje u tekstu: skup DE zapisa) i skup podataka o zastojsima u eksploataciji zrakoplova (nadalje u tekstu: skup OZ zapisa) biti objedinjen za sve sustave.

### 3.2.1. Podaci prikupljeni iz AHMS-a

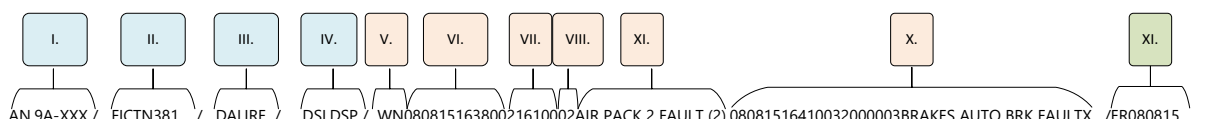
Kao što je navedeno u odjeljku 1.1, AHMS sustav isporučuje podatke o stanju zrakoplovnih sustava centralnom računalnom sustavu. Ovaj sustav zapravo gradi početnu dijagnozu o nepravilnostima u radu zrakoplovnih sustava te upozorava pilote i pohranjuje podatke na PFR-u, koje tehničko osoblje može naknadno analizirati.

Tijekom navedenog razdoblja prikupljanja podataka, na 422 leta zabilježene su poruke upozorenja za četiri zrakoplova tipa Airbus A319/20. Bitno je naglasiti da je ukupni broj letova za ova četiri zrakoplova u navedenom razdoblju bio veći, no na samo 422 leta zabilježeni su zapisi poruka upozorenja. Primjer pohranjivanja poruka i njihove strukture je vidljiv na slici 3-1.



Slika 3-1. Prikaz izvornog zapisa poruka upozorenja prikupljenih s PFR-a

Nestrukturirani zapis poruka upozorenja s PFR-a se može podijeliti u tri dijela. U prvom dijelu se sprema zapis o općim informacijama vezanim uz vrijeme i datum pristizanja serija poruka upozorenja, zatim se bilježi jedinstveni broj leta i registracija zrakoplova. Nakon toga slijedi dio s ključnim informacijama o svakoj poruci upozorenja. Detaljan prikaz zapisa tih informacija je dan na slici 3-2. Naposljetku, u trećem dijelu zapisa registrira se šifra računala i ID senzora koji su zabilježili poruku upozorenja.



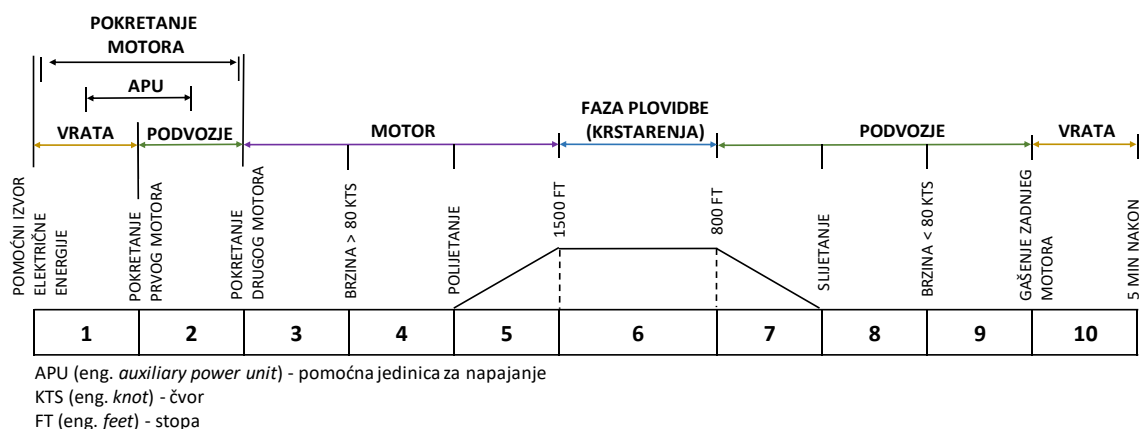
Slika 3-2. Prikaz logičkih cjelina jednog zapisa poruke upozorenja

Uvidom u strukturu zapisa ključnih informacija poruka upozorenja, vidljivo je kako se te informacije mogu razdijeliti u logičke cjeline koje će poslužiti kao ulazne značajke u daljnjoj analizi WN zapisa (tablica 3-2).

Tablica 3-2. Objašnjenje strukture poruke upozorenja

Enumeracija	Oznaka	Opis	Interpretacija
I.	AN.9A-XXX	Registracija zrakoplova	9A je kôd za Hrvatsku; XXX – registracija zrakoplova
II.	FICTN381	Broj leta	Svaki let ima jedinstveni broj (šifru) leta
III.	DALIRF	Odlazni aerodrom	Rim
IV.	DSL DSP	Dolazni aerodrom	Split
V.	WN	Kratice za upozorenje	Početak WN poruke
VI.	080815163800	Točan datum i vrijeme poruke upozorenja	Poruka snimljena 15. kolovoza 2008. u 16 sati i 38 minuta
VII.	216100	ATA poglavlje, podpoglavlje i podpodpoglavlje	21 – sustav grijanja, provjetravanja i klimatizacije zrakoplova; 61 – podsustav za regulaciju i indikaciju temperature; 00 – podpodsustav nije zabilježen
VIII.	02	Specifična faza leta	Pokretanje drugog motora
IX.	AIR PACK 2 FAULT (2)	Točna poruka upozorenja	Poruka upozorenja daje indikaciju da položaj ventila na spremniku ne odgovara odabranom položaju ili da ventil nije zatvoren. Broj dva u zagradama (2) znači da se ova WN poruka pojavila dva puta u fazi leta 02
X.	0808151641003200 0003BRAKES AUTO BRK FAULTX	Druga poruka upozorenja	Ova WN poruka pojavila se na ATA-i 32 (sustav podvozja) u fazi leta 03
XI.	FR	Lokacija s koje je prikupljena poruka upozorenja (eng. <i>failure report</i> , FR)	Točna šifra računala i senzora koji je zabilježilo određenu WN poruku

Početnom analizom nestrukturiranih zapisa poruka upozorenja uočeno je da se određena poruka upozorenja može pojaviti više puta u pojedinoj fazi leta. To je ključna informacija prilikom faze modeliranja jer se očekuje da postoji minimalni broj poruka upozorenja koji se treba pojaviti da bi se unio zapis u TLB. Drugim riječima, osim zapisa točne poruke upozorenja, potrebno je izdvojiti i točan broj ponavljanja poruke upozorenja. Analizom je također uočeno da se ista poruka upozorenja može pojaviti u različitim fazama leta. Na slici 3-3 prikazane su uobičajene faze leta za flotu civilnih zrakoplova za koju su prikupljeni podaci u ovom istraživanju (A320 flota).



Slika 3-3. Prikaz faza leta zrakoplova [95]

Iz slike 3-3 je vidljivo kako je let podijeljen u deset faza koje odgovaraju promjenama u konfiguraciji zrakoplova. Sukladno promjenama u konfiguraciji zrakoplova, u različitim fazama leta se prikazuju i različite poruke upozorenja [95]. U kritičnim fazama leta (3, 4, 5, 7, 8, 9) poruke upozorenja nisu vidljive pilotima na ECAM ekranu, no bilježe se na PFR-u.

Temeljem prošlih istraživanja prikazanih u odjeljku 1.2, uočeno je kako istraživači nisu razmatrali analizu poruku upozorenja po fazama leta. Iz tog razloga će u ovoj disertaciji ulazna značajka biti kombinacija poruka upozorenja po fazama leta.

### 3.2.2. Podaci o prošlim kvarovima ili nepravilnostima u radu sustava

Podaci o prošlim kvarovima prikupljeni su iz zasebne relacijske baze podataka koja preko jedinstvene šifre zapisa pruža detaljne informacije o zabilježenoj nepravilnosti i/ili kvaru. Prikaz značajki koje opisuju jedan zapis o kvaru je dan u tablici 3-3.

Tablica 3-3. *Prikaz značajki koje opisuju zapis o nepravilnosti/kvaru*

Izvorni naziv značajke	Opis
<i>Acreg</i>	Registracija zrakoplova
<i>Actype</i>	Tip zrakoplova
<i>Date</i>	Datum kada je zapis o nepravilnosti/kvaru registriran
<i>ATA</i>	ATA poglavlje za koje je zapis registriran
<i>Description</i>	Opis zapisa nepravilnosti/kvara
<i>Origin</i>	Izvor zapisa {PI; MA; QA}
<i>Flight</i>	Broj leta na kojemu je nepravilnost/kvar registriran
<i>Sta</i>	Stanica na kojoj je zapis stvoren (dolazni aerodrom)
<i>Log#</i>	Šifra zapisa iz TLB-a na temelju kojeg je kvar unesen u bazu
<i>Acc.Date</i>	Datum kada je nepravilnost/kvar otklonjen
<i>Csta</i>	Stanica na kojoj je kvar zatvoren
<i>Action</i>	Opis odrađenih akcija održavanja
<i>C</i>	MEL kategorija

Osim prethodno navedenih značajki prikazanih u tablici 3-3, u bazi se također nalaze i značajke vezane uz ljudske resurse: *UsedMh* – broj radnih sati potrebnih za ispravak kvara; *Skill* – kvalifikacija zrakoplovnog tehničara koji je odradio zahvat održavanja; *Mechanic* - šifra osobe koja je obavila aktivnost održavanja, *Inspector* – šifra osobe koja je provjerila/nadgledala odrađenu aktivnost održavanja. Budući da planiranje ljudskih i materijalnih resursa nije problematika ove disertacije, navedene značajke, kao i značajke *Sta* i *Csta*, nisu korištene u daljnjoj analizi.

Značajke koje će se koristiti u daljnjoj analizi su *Acreg*, *Actype*, *Date*, *ATA*, *Description*, *Origin*, *Flight*, *Acc.Date*, *Action*.

Kao što se može izlučiti iz tablice 3-3, *Origin* značajka može poprimiti tri vrijednosti – PI, MA, QA. Svaka od tih vrijednosti opisuje izvor zapisa o kvaru, tj., PI (eng. *pilot*) znači da je zapis nastao od strane pilota, MA (eng. *maintenance*) znači da je zapis nastao od strane osoblja za održavanje, dok QA (eng. *quality assurance*) znači da je zapis nastao temeljem zahtjeva odjela

Osiguranja kvalitete i sigurnosti. *Origin* značajka je važna jer predstavlja zapise o nepravilnostima/kvarovima koji su zahtijevali korektivne akcije održavanja. Iz tog razloga će se ta značajka, kao i značajke *Acreg* i *Actype*, koristiti u početnoj fazi filtriranja podataka, kako bi se odabrali samo oni zapisi o kvarovima koji su zabilježeni na Airbus floti zrakoplova i registrirani od strane pilota (PI-DE).

Značajke *Date*, *ATA* i *Flight* će biti ključne kod sastavljanja skupova podataka i povezivanja sa skupom podataka koji sadržava poruke upozorenja, kao i sa skupom podataka o zastojsima u eksploataciji zrakoplova. Također, preko značajki *Date* i *Acc. Date* će se provesti analiza o odgođenim popravcima i na taj način dodatno pročitati zapisi, što će biti detaljno objašnjeno u odjeljku 3.2.1.

Značajka *Description* će biti korištena za dubinsku analizu teksta i za predviđanje ishoda da li bi zrakoplov trebao biti prizemljen ili će let kasniti/biti otkazan zbog prirode opisa kvara.

### 3.2.3. Podaci o zastojsima u eksploataciji zrakoplova

Zrakoplovne vlasti i proizvođač zrakoplova zahtijevaju isporuku izvješća o pouzdanosti zrakoplovne flote od svakog zrakoplovnog prijevoznika. Za europske zemlje nadležno regulatorno tijelo je Europska agencija za sigurnost zračnog prometa (eng. *European Aviation Safety Agency*, EASA).

Izvješća o pouzdanosti zrakoplovne flote se smatraju radnim alatom za odjel Inženjeringa i tehničke podrške i odjel Održavanja s ciljem potvrđivanja ili opovrgavanja učinkovitosti programa održavanja zrakoplova (eng. *Aircraft Maintenance Program*, AMP) kroz statističku analizu različitih pokazatelja učinkovitosti kao što su vremena leta, vremena kašnjenja, itd. [96].

Svi prijevoznici isporučuju izvještaje agenciji na mjesečnoj razini za svoju flotu zrakoplova. Izvještaji se međusobno razlikuju, ovisno o zrakoplovnoj floti, broju zrakoplova, itd. No, u svakom izvještaju postoje podaci, tj. pokazatelji učinkovitosti, kao što su sati naleta svakog zrakoplova, broj prijavljenih kvarova i nesreća, planiranih zastoja, kašnjenja, otkazivanja i preusmjeravanja letova.

Pokazatelji pouzdanosti mogu se podijeliti u tri skupine. Prva skupina sadržava pokazatelje koji opisuju cijeli zrakoplov i osoblje povezano sa zrakoplovom; pilotsku posadu i tehničko osoblje za plovidbenost i održavanje zrakoplova [97].

U ovoj skupini mogu se naći pokazatelji kao što su: pilotski izvještaj o nepravilnostima na zrakoplovu (eng. *Pilot's Reports Rate*, PRR), broj prijava nepravilnosti od strane osoblja za održavanje (eng. *Maintenance Reports Rate*), pouzdanost otpreme (eng. *Dispatch Reliability*, DR), broj komponenata koje su trebale radioničko održavanje (eng. *Shop Rate*, SR), operacijski zastoji ili podaci o zastojima u eksploataciji zrakoplova (eng. *Operational Interruptions*), tehnički incidenti (eng. *Technical Incidents*, TI) [97].

Druga skupina sadržava pokazatelje koji preciznije opisuju zrakoplov i njegove sustave. Svaki pokazatelj u ovoj skupini izračunava se zasebno za svako ATA 100 poglavlje ili, u slučaju kvarova / uklanjanja / radioničkog popravka čak i za određenu komponentu. U ovoj skupini mogu se pronaći sljedeći pokazatelji: broj neplaniranih zamjena komponente (eng. *Removals Rate*, RR), intenzitet otkaza (eng. *Failure Rate*), srednje vrijeme između zamjena komponente (eng. *Mean Time Between Removals*, MTBR), srednje vrijeme između neplaniranih zamjena komponente (eng. *Mean Time Between Unscheduled Removals*, MTBUR) [97].

Treća skupina pokazatelja je vezana uz pogonske sustave. Budući da se radi o važnim sustavima zrakoplova, za njih se zasebno bilježe pokazatelji učinkovitosti. Neki od tih pokazatelja su: broj gašenja motora u toku leta (eng. *In-Flight Shut Down Rate*, IFSD), broj neplaniranih zamjena pogonskog sustava (eng. *Unscheduled Removals Rate*, URR) i broj pogonskih sustava koji su zahtijevali radioničko održavanje (eng. *Shop Visit Rate*, SVR) [97].

Iz prethodno navedenih skupina pokazatelja, za daljnju analizu će se koristiti podaci iz prve skupine pokazatelja. Točnije, analizirat će se podaci o zastojima u eksploataciji zrakoplova (kašnjenje i otkazivanje letova ili prizemljenje zrakoplova). Prikaz značajki koje opisuju svaki taj zapis se nalazi u tablici 3-4.

Tablica 3-4. Prikaz značajki koje opisuju podatke o zastojima u eksploataciji zrakoplova

Izvorni naziv značajke	Opis
<i>Date</i>	Datum kada je zapis o operacijskom zastoju registriran
<i>Event</i>	Događaj {Kašnjenje leta; Otkaz leta; Prizemljenje zrakoplova}
<i>A/C</i>	Registracija zrakoplova
<i>Description</i>	Opis zapisa o operacijskom zastoju registriran
<i>Corrective Actions</i>	Opis odrađenih akcija održavanja

Izvorni naziv značajke	Opis
<i>ATA</i>	ATA poglavlje za koje je zapis registriran
<i>Duration</i>	Trajanje popravka
<i>Station</i>	Stanica na kojoj je kvar otklonjen
<i>Code</i>	IATA kôd*

\*kôdovi koji opisuju tehničke razloge kašnjenja

Navedeni zapisi proizlaze iz kvarova koji se svakodnevno bilježe tijekom operacije leta ili na linijskom održavanju. Svaki izvještaj prikazuje uzrok kašnjenja i jasno identificira sustav ili komponentu na kojoj je došlo do kvara. Točan uzrok zastoja opisan je značajkom *Code* unutar koje se unose kôdovi, definirani od strane IATA-e, koji opisuju tehničke razloge kašnjenja. Na primjer, kôd 41 opisuje kvar zrakoplova kao uzrok operacijskih zastoja, kôd 42 podrazumijeva planirano održavanje, kôd 43 podrazumijeva neplanirano održavanje kao uzrok operacijskih zastoja, itd.

### 3.3. Strukturiranje i udruživanje skupova podataka

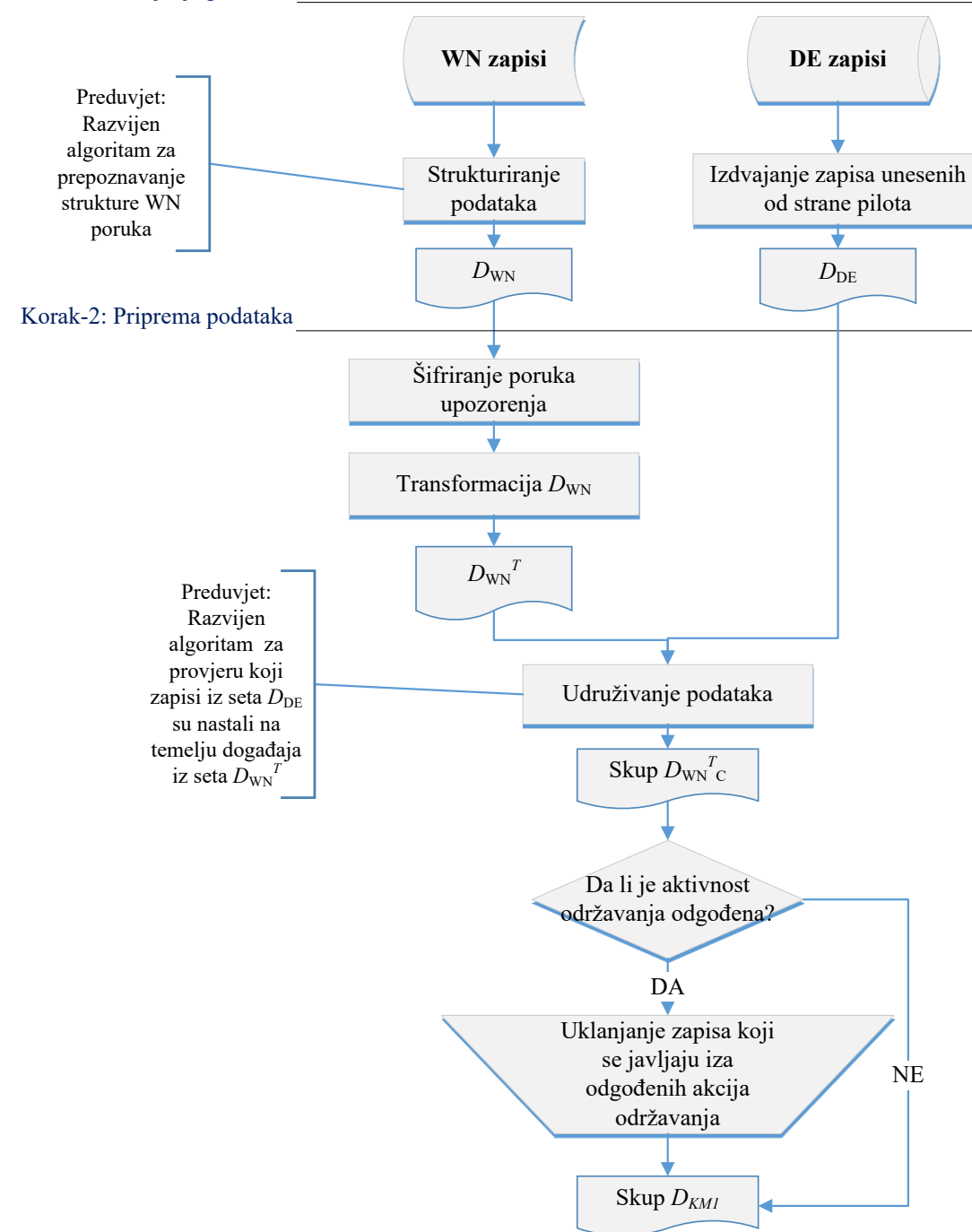
Kao što je opisano u odjeljku 3.1.1, skup WN zapisa se nalazi u obliku koji nije pogodan za analizu i integraciju s preostalim skupovima podataka jer njegova struktura nije u tabličnom formatu. Taj problem je moguće riješiti razvojem algoritama koji će prepoznati strukturu poruka upozorenja i transformirati je u tablicu. Detaljno razrađen postupak strukturiranja WN skupa podataka i integracija tog skupa i preostalih skupova podataka će biti prikazana u daljnjem tekstu.

#### 3.3.1. Postupak strukturiranja i integracija skupa WN i PI-DE

Postupak strukturiranja i integracije skupova podataka WN i DE proveden je na razini svakog zrakoplova  $Ac_i$  s ciljem stvaranja skupa  $D_{KMI}$  (slika 3-4). Razlog provođenja ovog postupka na razini zrakoplova je taj što se podaci iz AHMS-a prikupljaju zasebno za svaki zrakoplov, objedinjujući time poruke upozorenja za sve ugrađene sustave  $S_i$  na zrakoplovu  $Ac_i$ . Shodno tome, u fazi izgradnje klasifikacijskih modela prve grupe, ti skupovi podataka će se razdijeliti za svaki analizirani sustav  $S_i$ , što će biti prikazano u poglavlju 4.



### Korak-1: Izdvajanje podataka



Slika 3-4. Dijagram toka postupka za strukturiranje i udruživanje skupova WN i PI-DE

**Korak-1: Izdvajanje podataka.** Prva aktivnost se sastoji od strukturiranja poruka upozorenja koje su prikupljene u .txt datoteci za svaki zrakoplov  $Ac_i$ , gdje je  $i=1, \dots, 4$ . Kako bi se prikupljeni skup podataka, koji sadrži poruke upozorenja, mogao koristiti u daljnjoj analizi, razvijen je algoritam za izdvajanje poruka upozorenja.

Navedeni algoritam nije implementiran unutar platforme za dubinsku analizu podataka, već je razvijen primjenom VBA programskog jezika unutar softverskog alata MS Access. U prilogu

A nalazi se izvorni kôd algoritma za izdvajanje poruka upozorenja, a pseudo kôd je prikazan u tablici 3-5.

Tablica 3-5. Pseudo kôd algoritma za strukturiranje WN poruka

<b>Ulaz:</b>	.txt datoteka // datoteka koja sadržava nestrukturirane zapise poruka upozorenja (v. sliku 3-1)
<b>Izlaz:</b>	tablica // tablica koja sadržava strukturirane zapise poruka upozorenja (v. sliku 3-5)
1	<b>Stvori</b> varijable msgbox_odgovor, TextLine, pocetna_sekvenca, baza, rs, linija // deklaracija novih varijabli
2	<b>Pokreni</b> SQL naredbu za brisanje trenutnih zapisa u izlaznoj tablici // tablica T_Zrakoplov1
3	<b>Otvori</b> bazu i zapis tablice // kôd pridružen varijablama baza i rs
4	<b>Otvori</b> .txt datoteku // datoteka ZRAKOPLOV1.txt
5	<b>Sve dok nije kraj dokumenta</b> // petlja za prepoznavanje zapisa WN poruka u .txt datoteci
6	<b>ako</b> pocetna_sekvenca ← "- MPF/" <b>onda</b> // početak zapisa
7	<b>stvori</b> varijable registracija, broj_leta, datum, vrijeme, odlazni_aerodrom, WN_poruka, date_warning_poruka, time_wn_poruka, ATA, POD_ATA, POD_POD_ATA, faza_leta, zabiljezana_poruka, tocan_br_ponavljanja_WNa
8	<b>ako</b> warning_poruka ← "WN" <b>onda</b> //provjeri da li je pronađena WN poruka
9	<b>spremi</b> zapise u bazu podataka // dio linije teksta koji je spremljen unutar varijabli je pridodan kao vrijednost određene značajke u tablici T_Zrakoplov1
10	<b>kraj</b>
11	<b>kraj</b>
12	<b>pokreni</b> petlju
13	<b>Zatvori</b> datoteku // datoteka ZRAKOPLOV1.txt
14	<b>Zatvori</b> tablicu // svi zapisi su dodani u tablicu T_Zrakoplov1

Zadatak ovog algoritma je da uzme ulaznu datoteku u izvornom zapisu te da, sukladno prikazu logičkih cjelina poruka upozorenja iznesenom u odjeljku 3.1.1 na slici 3-1, zabilježi svaku cjelinu kao zasebnu značajku i stvori izlaznu datoteku u tabličnom obliku (slika 3-5). Drugim riječima, rezultat izlazne datoteke je skup podataka  $D_{WN}$  koji se sastoji od značajki  $A_{WN} = \{registracija, broj\_leta, datum, vrijeme, odlazni\_aerodrom, dolazni\_aerodrom, warning\_poruka, date\_warning\_poruka, time\_warning\_poruka, ATA, POD\_ATA, POD\_POD\_ATA, faza\_leta, zabiljezana\_poruka, tocan\_br\_ponavljanja\_WNa\}$  i broja uzoraka  $d_{WNj}$ , gdje je  $j=1, \dots, m$ , a  $m$  se razlikuje za svaki zrakoplov  $Ac_i$ .

- MPF/AN.9A-XXX/FICTN381 /DM080815172200/DALIRF/DSL DSP/WN08081516380021610002AIR PACK 2 FAULT (2),08081516410032000003BRAKES AUTO BRK FAULTX														
Registracija	Broj_leta	Datum	Vrijeme	Odlazni_aerodrom	Dolazni_aerodrom	Wn_poruka	Date_warning_poruka	Time_wn_poruka	ATA	POD_	POD_POD_ATA	Faza_leta	Zabiljezana_poruka	Tocan_br_ponavljanja_WNa
9A-XXX	FICTN381	15.08.2008	17:22	DALIRF	DSL DSP	WN	15.08.2008	16:38	21	61	00	02	AIR PACK 2 FAULT	2
9A-XXX	FICTN381	15.08.2008	17:22	DALIRF	DSL DSP	WN	15.08.2008	16:41	32	0	00	03	BRAKES AUTO B	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Slika 3-5. Prikaz značajki skupa  $D_{WN}$

Druga aktivnost obuhvaća izdvajanje zapisa o kvarovima iz relacijske baze koja sadržava povijesne zapise o kvarovima za svaki zrakoplov  $Ac_i$ . Primjenom strukturnog upitnog jezika (eng. *structured query language*, SQL) filtrirani su samo onih zapisi koji su registrirani od strane pilota (PI-DE) i na Airbus floti zrakoplova (A2) te je stvoren skup  $D_{DE}$ . Skup podataka  $D_{DE}$  sastoji se od značajki  $A_{DE} = \{A_{creg}, A_{ctype}, Date, ATA, Description, Origin, Flight, Acc.Date, Action\}$  i broja uzoraka  $d_{DEj}$ , gdje je  $j=1, \dots, m$ , a  $m$  se razlikuje za svaki zrakoplov  $Ac_i$  (slika 3-6).

<i>Acreg</i>	<i>Actype</i>	<i>Date</i>	<i>ATA</i>	<i>Description</i>	<i>Origin</i>	<i>Acc.Date</i>	<i>Action</i>
9A-XXX	A2	15.08.2008	22	FMGC #2 U/S PART	PI	24.08.2008	FMGC #2 REPL ACC.AMM
9A-XXX	A2	17.09.2008	23	PRAM INOP NO SP	PI	06.10.2008	TAPE REPRODUCER ANNC
...	...	...	...	...	...	...	...
9A-XXX	A2	13.10.2008	21	PACK 1 OVERHEAT	PI	14.10.2008	SENSOR 15HH REPLACED.
9A-XXX	A2	20.10.2008	31	MAINTENANCE ST	PI	09.11.2008	
...	...	...	...	...	...	...	...

Slika 3-6. Prikaz značajki skupa  $D_{DE}$

**Korak-2: Priprema podataka.** Analogno prethodnom koraku, priprema podataka s ciljem stvaranja ulaznog skupa podataka za učenje  $D_{KMI}$  bit će odrađena neovisno o platformi za dubinsku analizu podataka primjenom MS Excel alata.

Prvo, potrebno je transformirati skup podataka  $D_{WN}$  kako bi se moglo razlučiti koji zapisi iz skupa podataka  $D_{DE}$  su nastali na temelju događaja u skupu podataka  $D_{WN}$ . Značajka *zabilježena\_poruka*, koja sadržava tekstualni opis svake poruke upozorenja, pretvorena je u numerički zapis na način da svaka  $z_i$  vrijednost značajke *zabilježena\_poruka*, gdje je  $i=0, \dots, n$ , a  $n$  je različit za svaki sustav  $S_i$  i zrakoplov  $Ac_i$ , poprimi svoju šifru koja će biti spremljena unutar nove značajke *sifra\_WNa* s  $a_j$  vrijednosti, gdje je  $j=0, \dots, z_n$  (slika 3-7).

...	<i>ATA</i>	<i>POD_ATA</i>	<i>POD_POD_ATA</i>	<i>Faza_leta</i>	<i>Zabilježena_poruka</i>	<i>Tocan_br_ponavljanja_</i>	<i>sifra_WNa</i>
...	21	61	00	02	AIR PACK 2 FAULT	2	21-00
...	32	0	00	3	BRAKES AUTO BR	1	32-00
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

Slika 3-7. Prikaz dodjeljivanja šifre svakoj poruci upozorenja unutar skupa  $D_{WN}$

Nakon što su svi tekstualni opisi poruka upozorenja zamijenjeni jedinstvenim šiframa, unutar skupa  $D_{WN}$  stvorena je hibridna značajka *faza\_leta/sifra\_WNa* koja povezuje značajke *sifra\_WNa* i *faza\_leta* (slika 3-8).

...	ATA	POD_ATA	POD_POD_ATA	Faza_leta	Zabiljezana_poruka	Tocan_br_ponavljanja_WNa	sifra_WNa	faza_leta/sifra_WNa
...	21	61	00	02	AIR PACK 2 FAULT	2	21-00	02/21-00
...	32	0	00	03	BRAKES AUTO BR	1	32-00	03/32-00
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...

Slika 3-8. Prikaz stvaranja hibridne ulazne značajke unutar skupa  $D_{WN}$

Konačno, transformacija skupa podataka  $D_{WN}$  odrađena je primjenom modula *Pivot Table* unutar MS Excel alata. Skup podataka  $D_{WN}^T$  sastoji se od niza uzoraka  $d_{WN}^T$  i svaki uzorak je predstavljen kao  $n$ -dimenzionalni vektor  $X_{WN}^T = (x_{WN}^T_1, \dots, x_{WN}^T_n)$ , gdje su  $x_{WN}^T_1$  i  $x_{WN}^T_2$  vrijednosti značajki *datum* i *broj\_leta*, a  $x_{WN}^T_3$  do  $x_{WN}^T_{n-1}$  vrijednosti broja ponavljanja značajke *faza\_leta/sifra\_WNa* za svaki sustav  $S_i$  i zrakoplov  $Ac_i$  (slika 3-9).

Datum	Broj_leta	02/21-00	03/21-01	04/21-02	...	$x_{WN}^T_{n-1}$
02.09.2008	(OU660)	0	0	0	...	0
22.09.2008	(CTN656)	0	0	0	...	0
28.09.2008	(CTN410)	0	0	0	...	0
04.10.2008	(CTN5H)	0	0	0	...	0
16.10.2008	(CTN449)	0	0	0	...	0
26.11.2008	(CTN493)	0	0	0	...	0
15.12.2008	(CTN656)	0	0	0	...	0
26.12.2008	(CTN491)	0	0	0	...	0
28.12.2008	(CTN411)	0	0	0	...	0
06.01.2009	(CTN456)	0	0	0	...	0
07.01.2009	(CTN668)	0	0	0	...	0
10.01.2009	(CTN5N)	0	0	0	...	0
...	...	...	...	...	...	...

Slika 3-9. Prikaz transformacije skupa  $D_{WN}$  u skup  $D_{WN}^T$  primjenom modula *Pivot Table*

Budući da skupovi  $D_{DE}$  i  $D_{WN}$  nisu izravno povezani, temeljem iskustvenih spoznaja prikupljenih od strane zračnog prijevoznika, može se pretpostaviti da je svaki povijesni (pilotski) zapis o nepravilnosti u radu ili kvaru sustava iz skupa  $D_{DE}$  nastao na temelju vektora  $X_{WN}^T$  iz skupa  $D_{WN}^T$  ako i samo ako je značajka *datum* iz skupa  $D_{WN}^T$  jednaka značajki *Date* iz skupa  $D_{DE}$  i značajka *ATA* iz skupa  $D_{WN}^T$  jednaka značajki *ATA* iz skupa  $D_{DE}$ .

Iz tog razloga razvijen je algoritam koji omogućava prepoznavanje zapisa iz skupa  $D_{DE}$  nastalih na temelju vektora  $X_{WN}^T$  iz skupa  $D_{WN}^T$ . Izvorni kôd algoritma se nalazi u prilogu B.

Nakon što su prepoznati uzorci iz skupa  $D_{WN}^T$  koji su uzrokovali uzorke u skupu  $D_{DE}$ , unutar skupa  $D_{WN}^T$  stvorena je značajka *PI-zapis?* koja predstavlja klasu  $C_m$ , gdje je  $m=2$ , i time je stvoren novi skup  $D_{WN}^T_C$ , gdje je broj uzoraka  $d_{WN}^T_C = d_{WN}^T \neq d_{DEj}$ .

Prethodno opisane tvrdnje mogu se i matematički izraziti:

$$\text{akko } \{datum, ATA\} \in D_{WN}^T = \{Date, ATA\} \in D_{DE} \Rightarrow \{C_m\} \in D_{WN}^T C = 1$$

Nadalje, dodavanjem značajke *Acc.Date* iz skupa  $D_{DE}$  pored svakog uzorka  $d_{WN}^T C$  gdje je vrijednost  $C = 1$ , stvorena je nova značajka pod nazivom *DTTR* (eng. (i dalje) *delay time to repair*). Ukoliko je vrijednost značajke  $DTTR > 0$ , smatra se da je aktivnost održavanja odgođena te svi  $d_{WN}^T C$  uzorci koji se javljaju u periodu odgode mogu biti uklonjeni iz skupa  $D_{WN}^T C$  (slika 3-10).

<i>Datum</i>	<i>Broj_leta</i>	02/21-00	03/21-01	04/21-02	...	$x_{WN}^T_{n-1}$	<i>PI-zapis? [1 ili 0]</i>	<i>Acc.Date</i>	<i>DTTR</i>
02.09.2008	(OU660)	0	0	0	...	0	0		
22.09.2008	(CTN656)	0	0	0	...	0	0		
28.09.2008	(CTN410)	0	0	0	...	0	0		
04.10.2008	(CTN5H)	0	0	0	...	0	0		
16.10.2008	(CTN449)	0	0	0	...	0	0		
26.11.2008	(CTN493)	0	0	0	...	0	0		
15.12.2008	(CTN656)	0	0	0	...	0	0		
26.12.2008	(CTN491)	0	0	0	...	0	0		
28.12.2008	(CTN411)	0	0	0	...	0	0		
<b>06.01.2009</b>	<b>(CTN456)</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>...</b>	<b>0</b>	<b>1</b>	<b>07.01.2009</b>	<b>1</b>
07.01.2009	(CTN668)	0	0	0	...	0	0		
10.01.2009	(CTN5N)	0	0	0	...	0	0		
...	...	...	...	...	...	...	...	...	...

Slika 3-10. Prikaz preprocesuiranja skupa  $D_{WN}^T C$

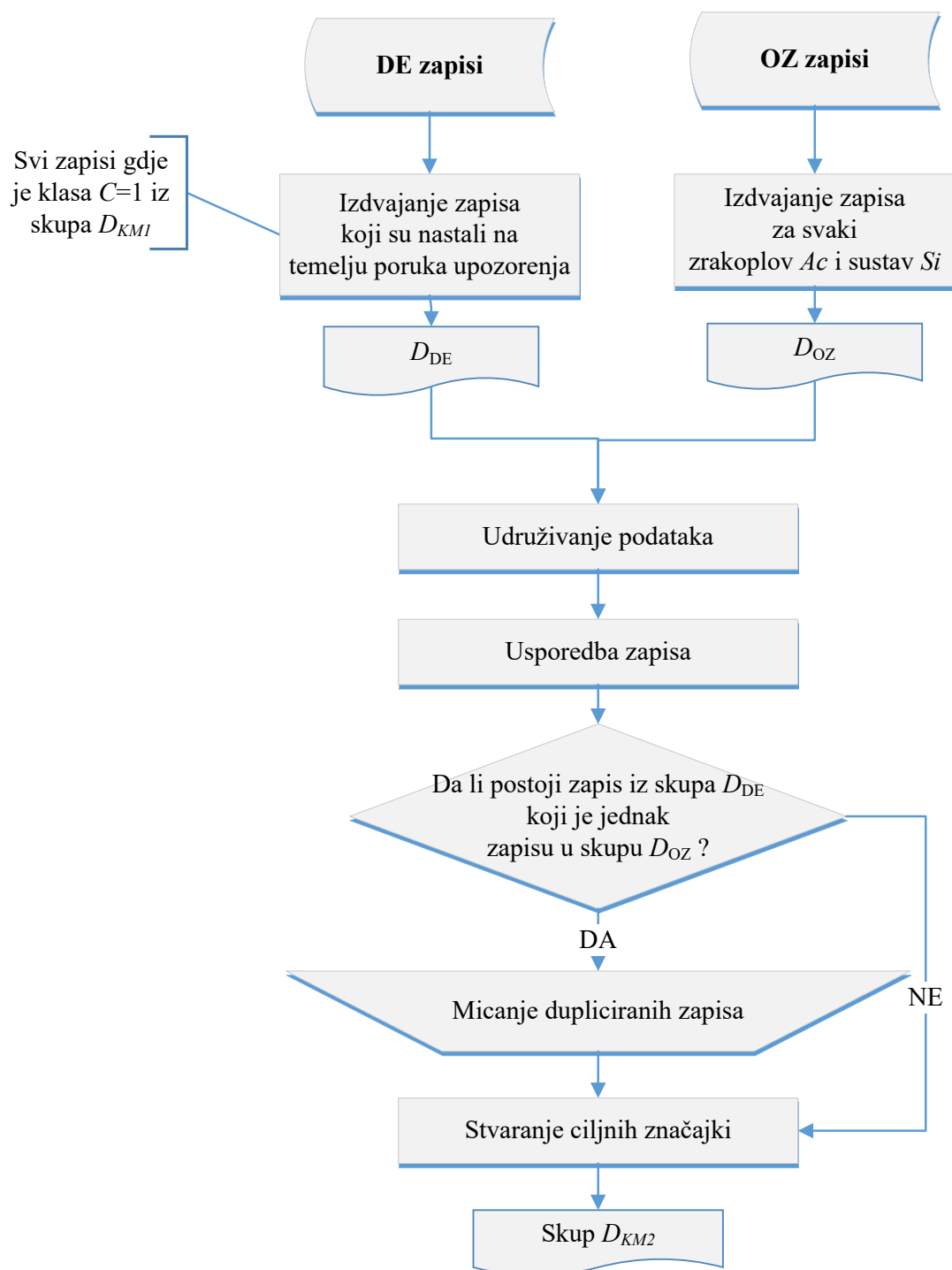
Naposljetku, krajnji rezultat koraka pripreme podataka je skup  $D_{KMI}$  koji će služiti u daljnjoj analizi za otkrivanje relevantnih značajki i izradu klasifikacijskih modela prve grupe (slika 3-11).

<i>Datum</i>	<i>Broj_leta</i>	02/21-00	03/21-01	04/21-02	...	$x_{WN}^T_{n-1}$	<i>PI-zapis? [1 ili 0]</i>	<i>Acc.Date</i>	<i>TTR</i>
02.09.2008	(OU660)	0	0	0	...	0	0		
22.09.2008	(CTN656)	0	0	0	...	0	0		
28.09.2008	(CTN410)	0	0	0	...	0	0		
04.10.2008	(CTN5H)	0	0	0	...	0	0		
16.10.2008	(CTN449)	0	0	0	...	0	0		
26.11.2008	(CTN493)	0	0	0	...	0	0		
15.12.2008	(CTN656)	0	0	0	...	0	0		
26.12.2008	(CTN491)	0	0	0	...	0	0		
28.12.2008	(CTN411)	0	0	0	...	0	0		
<b>06.01.2009</b>	<b>(CTN456)</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>...</b>	<b>0</b>	<b>1</b>	<b>07.01.2009</b>	<b>1</b>
10.01.2009	(CTN5N)	0	0	0	...	0	0		
...	...	...	...	...	...	...	...	...	...

Slika 3-11. Prikaz skupa podataka  $D_{KMI}$

### 3.3.2. Postupak integracije skupa PI-DE i skupa OZ

Sukladno postupku integracije podataka iz prethodnog odjeljka, postupak udruživanja skupova DE i OZ proveden je nezavisno o platformi za dubinsku analizu podataka primjenom MS Excel alata, kako je prikazano na slici 3-12.



Slika 3-12. Dijagram toka postupaka za udruživanje skupova PI-DE i OZ

Na temelju zapisa iz skupa podataka  $D_{KMI}$  izdvojeni su oni uzorci  $d_{KMIi}$  koji su svrstani u klasu  $C = 1$ . Za izdvojene uzorke iz izvornog seta zapisa o kvarovima  $D_{DE}$  izlučene su značajke  $A_{DE} = \{A_{reg}, Date, ATA, Description\}$ . Ti uzorci su sjedinjeni s uzorcima iz skupa  $D_{OZ}$  iz kojeg su, analogno prethodnom skupu, izlučene značajke  $A_{OZ} = \{A/C, Date, ATA, Description\}$ . Nakon sjedinjavanja skupova, provedena je usporedba zapisa i micanje dupliciranih zapisa. Naposljetku, stvorene su ciljane značajke *Utjecaj na raspoloživost* i *Posljedica* koje sadržavaju informacije o tome da li je opisani kvar utjecao na raspoloživost (da = 1, ne = 0) i koja je bila posljedica (zastoj, otkaz leta ili prizemljenje zrakoplova). Krajnji rezultat ovog postupka je skup  $D_{KM2}$  koji će služiti za izradu druge grupe klasifikacijskih modela (slika 3-13).

<i>Acreg</i>	<i>Date</i>	<i>ATA</i>	<i>Description</i>	<i>Origin</i>	<i>Utjecaj na raspoloživost</i>	<i>Posljedica</i>
9A-XXX	26.06.2008	27	CAPT. SIDE STI	PI	0	None
9A-XXX	09.02.2009	27	FUEL THE FUEL	PI	0	None
9A-XXX	12.03.2009	27	MAINTENANCE	PI	0	None
9A-XXX	19.07.2009	27	ELAC#1pitch fault	OZ	1	Delay
9A-XXX	14.09.2009	27	RH Aileron servoc	OZ	1	Delay
9A-XXX	14.09.2009	27	F/CTL AIR SERV	PI	0	None
9A-XXX	20.09.2009	27	MAINTENANCE	PI	0	None
9A-XXX	30.09.2009	27	MAINTENANCE	PI	0	None
....	....	....	....	....	....	....

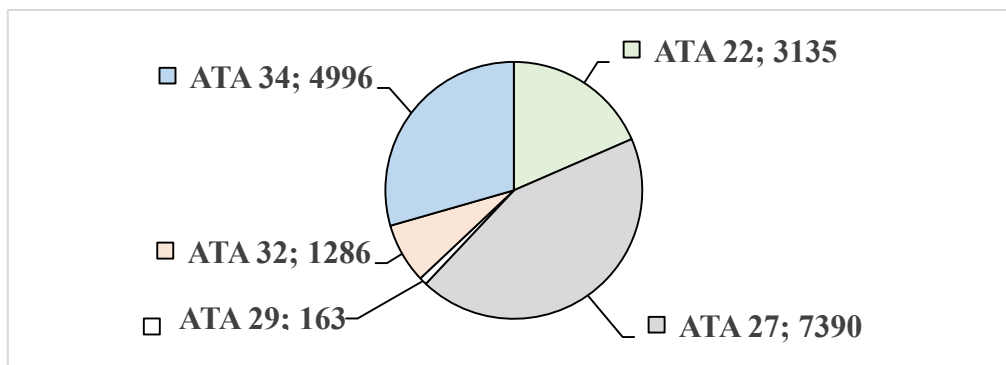
Slika 3-13. Prikaz dijela uzoraka iz skupa podataka  $D_{KM2}$  zrakoplova  $Ac_i$

### 3.4. Eksplorativna analiza podataka

U odjeljku 2.4 izvedeni su zaključci kako je u procesu dubinske analize podataka ključno razumijevanje skupova podataka s ciljem odabira klasifikacijskog algoritma i statističkih mjera za evaluaciju modela. S tom svrhom provedena je eksplorativna analiza podataka.

#### 3.4.1. Skup $WN$ i $DE$ zapisa

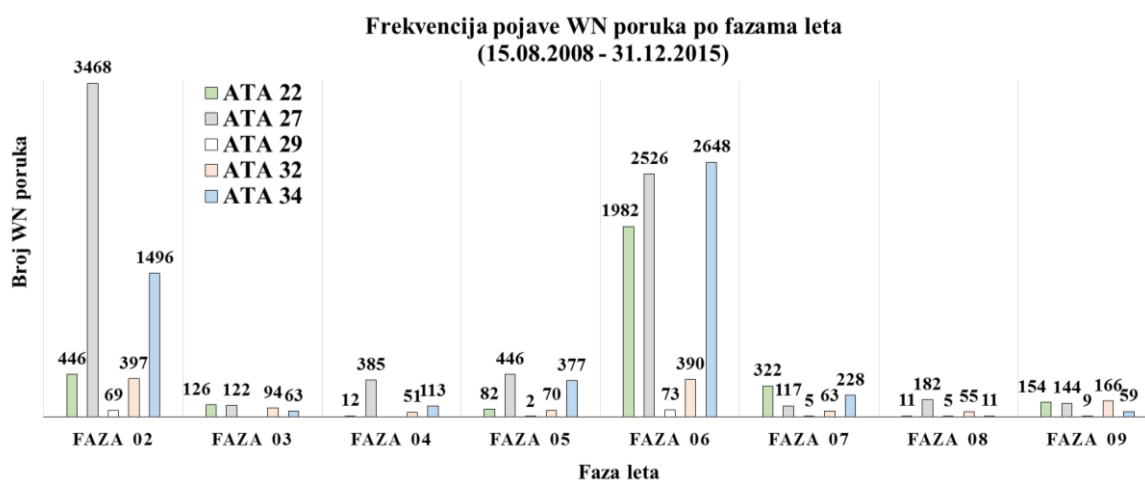
Nakon strukturiranja poruka upozorenja i sastavljanja skupa  $D_{WN}$ , analizom je otkriveno da je u periodu od 15.08.2008. do 31.12.2015. godine zabilježeno ukupno 16 970 poruka upozorenja po analiziranim sustavima na sva četiri zrakoplova (slika 3-14).



Slika 3-14. Prikaz broja poruka upozorenja po analiziranim sustavima na razini flote

Vidljivo je kako je najmanji broj poruka generirao sustav hidraulike (ATA 29), dok su ostali sustavi; ATA 22 (sustav automatskog leta), ATA 27 (sustav kontrole leta), ATA 32 (sustav podvozja) i ATA 34 (sustav navigacije) generirali znatno veći broj poruka ne temelju kojeg je moguće izgraditi klasifikacijske modele. Zbog malog broja uzorka poruka upozorenja koje je generirao sustav ATA 29 (samo 163 uzorka), na ovom skupu podataka provest će se samo analiza utjecajnih značajki i dubinska analiza teksta.

Budući da će ulazne značajke za izgradnju klasifikacijskog modela predstavljati kombinaciju poruke upozorenja i faze leta u kojoj se poruka pojavila, na slici 3-15 je prikazana frekvencija pojave poruka upozorenja po fazama leta za svaki od sustava koji će biti analiziran. Sve poruke upozorenja koje su generirane u kritičnim fazama leta (kao što su faze uzlijetanja i spuštanja), nisu vidljive pilotima na ECAM-u i prikazuju se u tek u kasnijim fazama leta. No, pojava tih poruka u točnim fazama leta je zabilježena na PFR-u i može poslužiti u daljnjoj analizi posadi za održavanje.



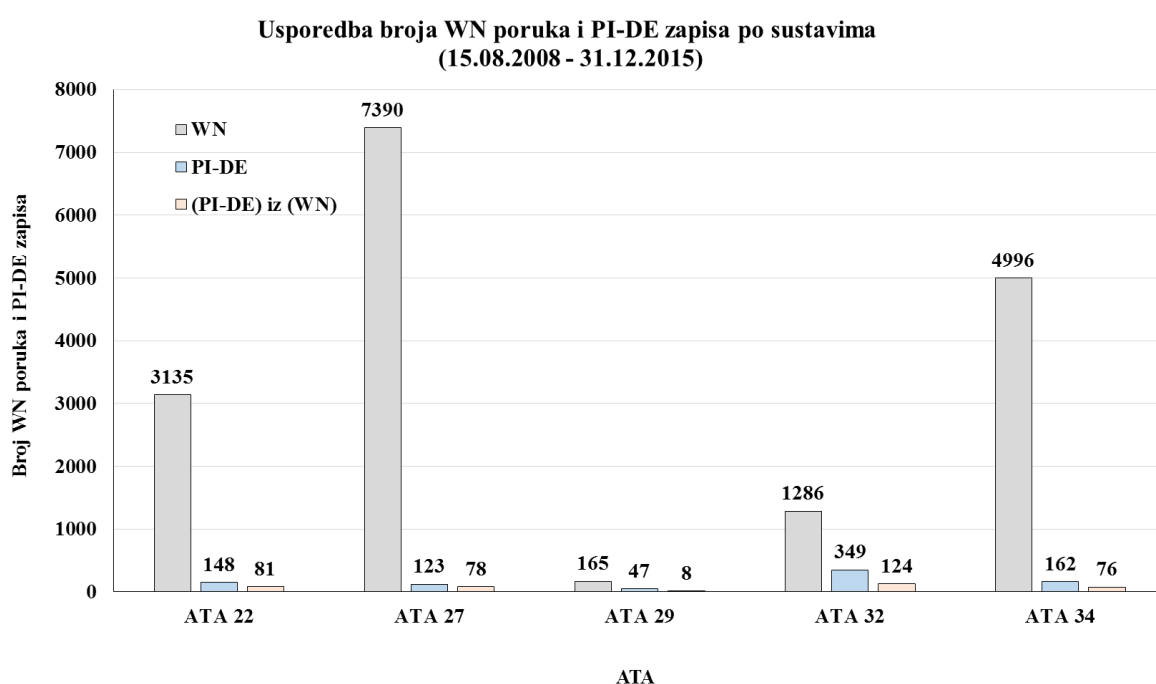
Slika 3-15. Prikaz frekvencije pojave poruka upozorenja po fazama leta za ATA 22, ATA 27, ATA 29, ATA 32 i ATA 34 sustave



Iz slike 3-15 je vidljivo kako su analizirani sustavi generirali najviše poruka upozorenja u fazi pokretanja motora (faza 02) i fazi plovidbe (faza 06). U fazi ubrzavanja najveći broj poruka upozorenja generirao je sustav ATA 27, dok je u fazi plovidbe najviše poruka generirao sustav ATA 34.

Nadalje, analizom skupa  $D_{WN}^T C$  (skup nastao udruživanjem skupa  $D_{DE}$  i  $D_{WN}$ ), otkriveno je da su 367 PI-DE zapisa nastala na temelju WN poruka na analiziranim sustavima.

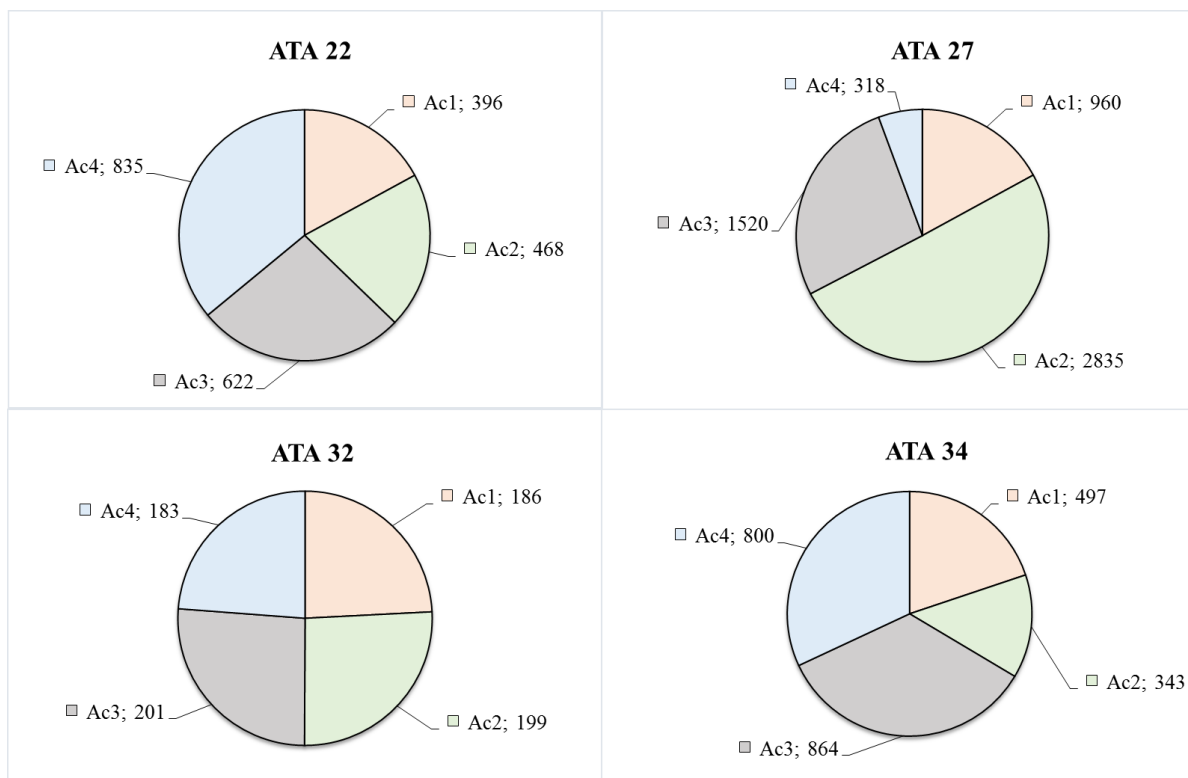
Usporedbom uzorak u  $D_{WN}^T C$  skupu podataka, na slici 3-16 prikazane su frekvencije pojave WN poruka i PI-DE zapisa nastalih temeljem tih poruka po svakom sustavu.



Slika 3-16. Usporedba broja generiranih WN poruka i PI-DE zapisa nastalih temeljem istih poruka po svakom sustavu

Iz grafa prikazanog na slici 3-16 vidljivo je da postoji neuravnoteženost skupova podataka. Drugim riječima, velik broj primjeraka poruka upozorenja neće rezultirati pilotskim zapisom u TLB i ciljna varijabla, koja predviđa pojavu zapisa, će sadržavati većinom klasu (razred)  $C = 0$ .

Analizom skupova  $D_{KMI}$  za svaki analizirani sustav uočen je broj uzoraka na razini zrakoplova  $Ac_i$  (slika 3-17).

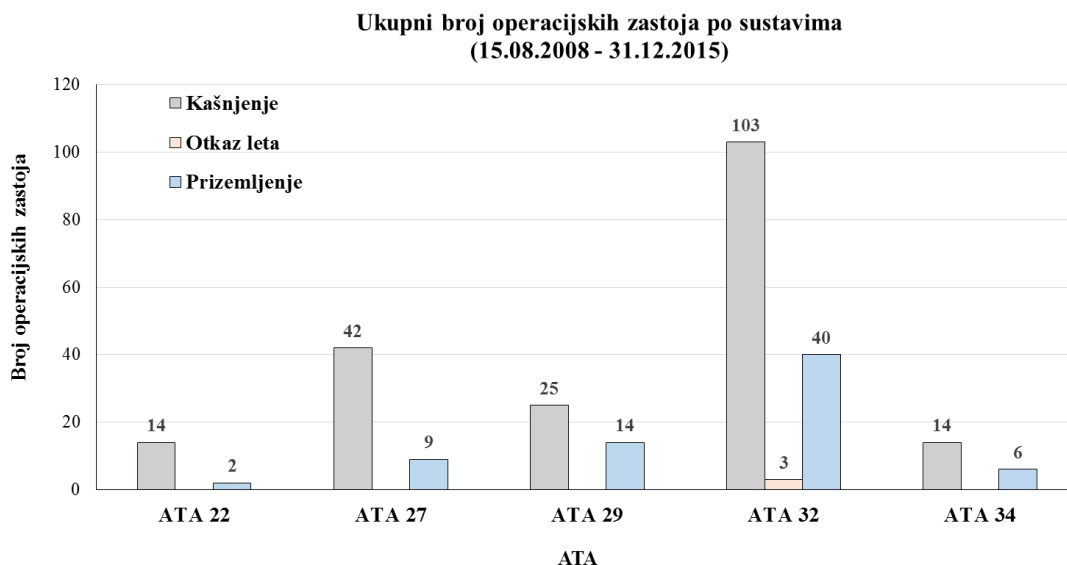


Slika 3-17. Prikaz broja uzoraka u skupu  $D_{KMI}$  za  $Ac_i$  zrakoplov i analizirani sustav

Na slici 3-17 nije prikazan sustav ATA 29 jer navedeni sustav nije generirao dovoljan broj poruka upozorenja za izgradnju skupa  $D_{KMI}$ . Analizirajući sustave ATA 22, ATA 27, ATA 32 i ATA 34, iz grafikona prikazanih na slici 3-17, vidljivo je kako se najveći  $D_{KMI}$  skup po zrakoplovu  $Ac_i$  sastoji od 2.835 uzorka (ATA 27, zrakoplov  $Ac_2$ ), a najmanji  $D_{KMI}$  skup od 183 uzorka (ATA 32, zrakoplov  $Ac_4$ ). Budući da većina  $D_{KMI}$  skupova podataka ima između sto i tisuću uzoraka, na temelju zaključka provedenih u odjeljku 2.3.5.1, prilikom izgradnje klasifikacijskog modela koristit će se  $k$ -struka unakrsna validacija.

### 3.4.2. Skup DE i OZ zapisa

Analizom skupa OZ klasificirani su događaji operacijskih zastoja (kašnjenje leta, otkaz leta ili prizemljenje) temeljem prirode opisa kvara, kao što je to prikazano na slici 3-18.



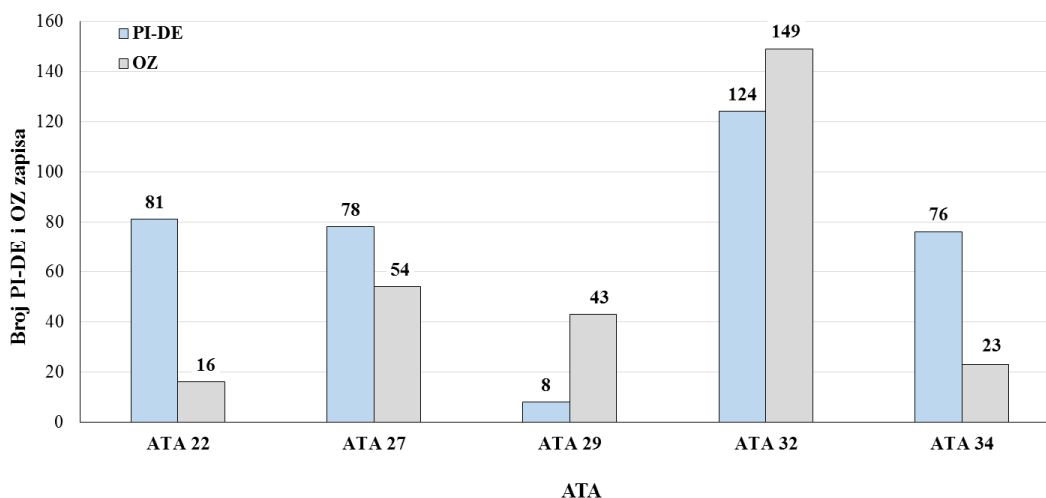
Slika 3-18. Klasifikacija operacijskih zastoja na razini analiziranih sustava

Uvidom u klasifikaciju događaja operacijskih zastoja prikazanih na slici 3-18 može se zaključiti da je najviše kvarova rezultiralo kašnjenjem leta i prizemljenjem zrakoplova (eng. *aircraft on ground*, AOG), a najmanje otkazom leta.

Budući da su prava znanja u ovom skupu podataka sakrivena unutar nestrukturiranih tekstualnih zapisa, klasifikacijski modeli druge grupe obuhvatit će dubinsku analizu podataka. No, kako bi se dodatno analizirali i nestrukturirani PI-DE zapisi nastali na temelju WN poruka, ti zapisi će biti preklapljeni s OZ zapisima, na način kako je to objašnjeno u odjeljku 3.3.2.

Slika 3-19 prikazuje broj PI-DE zapisa nastalih temeljem WN poruka u odnosu na broj OZ zapisa u analiziranom periodu po zrakoplovnim sustavima.

Usporedba broja PI-DE i OZ zapisa po sustavima  
(15.08.2008 - 31.12.2015)



Slika 3-19. Usporedba broja PI-DE zapisa i OZ zapisa po analiziranim sustavima

Usporedbom prethodno navedenih skupova, sa slike 3-19 može se zaključiti kako je većina zapisa o kvarovima, koji su rezultirali operacijskim zastoјima na ATA 29 i ATA 32 sustavu, nastala iz drugih izvora (npr. od strane održavanja, MA izvor), a ne isključivo temeljem PI-DE zapisa. Međutim, i te zapise je korisno analizirati kako bi se ubuduće olakšao proces identifikacije grešaka i planiranja zaliha na skladištu.

### 3.5. Zaključak poglavlja i smjernice za nastavak istraživanja

Temeljem analiza provedenih u odjeljku 3.4, smjernice za nastavak istraživanja su sljedeće:

1. Eksplorativnom analizom je utvrđena neuravnoteženost skupova podataka. Drugim riječima, manjinska klasa, koja je bitnija za domenski problem, ima mali broj uzoraka u skupu  $D_{KMI}$ . Kako bi se riješio problem neuravnoteženosti, prilikom izrade klasifikacijskih modela prve grupe, provest će se balansiranje podataka metodama uzorkovanja.
2. Shodno zaključcima iz odjeljka 2.3.5.2, zbog neuravnoteženosti skupova podataka, glavne mjere koje će se koristiti za evaluaciju modela su *F-mjera*, *osjetljivost* i *specifičnost*.
3. Budući da većina  $D_{KMI}$  skupova podataka ima između sto i tisuću uzoraka (slika 3-17), prilikom izgradnje klasifikacijskog modela koristit će se *k-struka unakrsna validacija*.
4. Klasifikacijski algoritam koji će se koristiti kod izgradnje modela će biti Naive Bayes (ili njegove modifikacije) jer je prednost tog algoritma što dobro predviđa i na malim i

velikim skupovima podataka za učenje. Druga važna prednost tog algoritma je ta što pretpostavlja da su ulazne značajke neovisne, što i je slučaj u skupu  $D_{KMI}$ .

5. Budući da je većina ljudskog znanja pohranjena u tekstualnom obliku, dubinska analiza zapisa o nepravilnostima u radu sustava, kvarova i operacijskih zastoja je potrebna kako bi se ubuduće olakšao proces identifikacije grešaka (eng. *troubleshooting*) i planiranja zaliha na skladištu.

## 4. FORMULACIJA PROCESA ZA OTKRIVANJE RELEVANTNIH ZNAČAJKI I IZGRADNJU KLASIFIKACIJSKIH MODELA

U ovome poglavlju prikazani su procesi za otkrivanje relevantnih značajki i izgradnju klasifikacijskih modela prve i druge grupe s ciljem ostvarivanja znanstvenih doprinosa #2 i #3. Procesu su razvijeni korištenjem RapidMiner platforme za dubinsku analizu podataka nad skupovima podataka  $D_{KMI}$  i  $D_{KM2}$ . Temeljem zaključaka i smjernica iz prethodnih poglavlja, unutar razvijenih procesa primijenjeni su odgovarajući operatori koji omogućavaju učitavanje skupova podataka, uzorkovanje podataka, pretvaranje podataka iz numeričkih u kategoričke, primjenu različitih klasifikacijskih algoritama, evaluaciju modela, itd. Prikaz povezanih operatora unutar razvijenih procesa u ovoj disertaciji se nalazi u prilogu C.

### 4.1. Formulacija procesa za otkrivanje relevantnih značajki

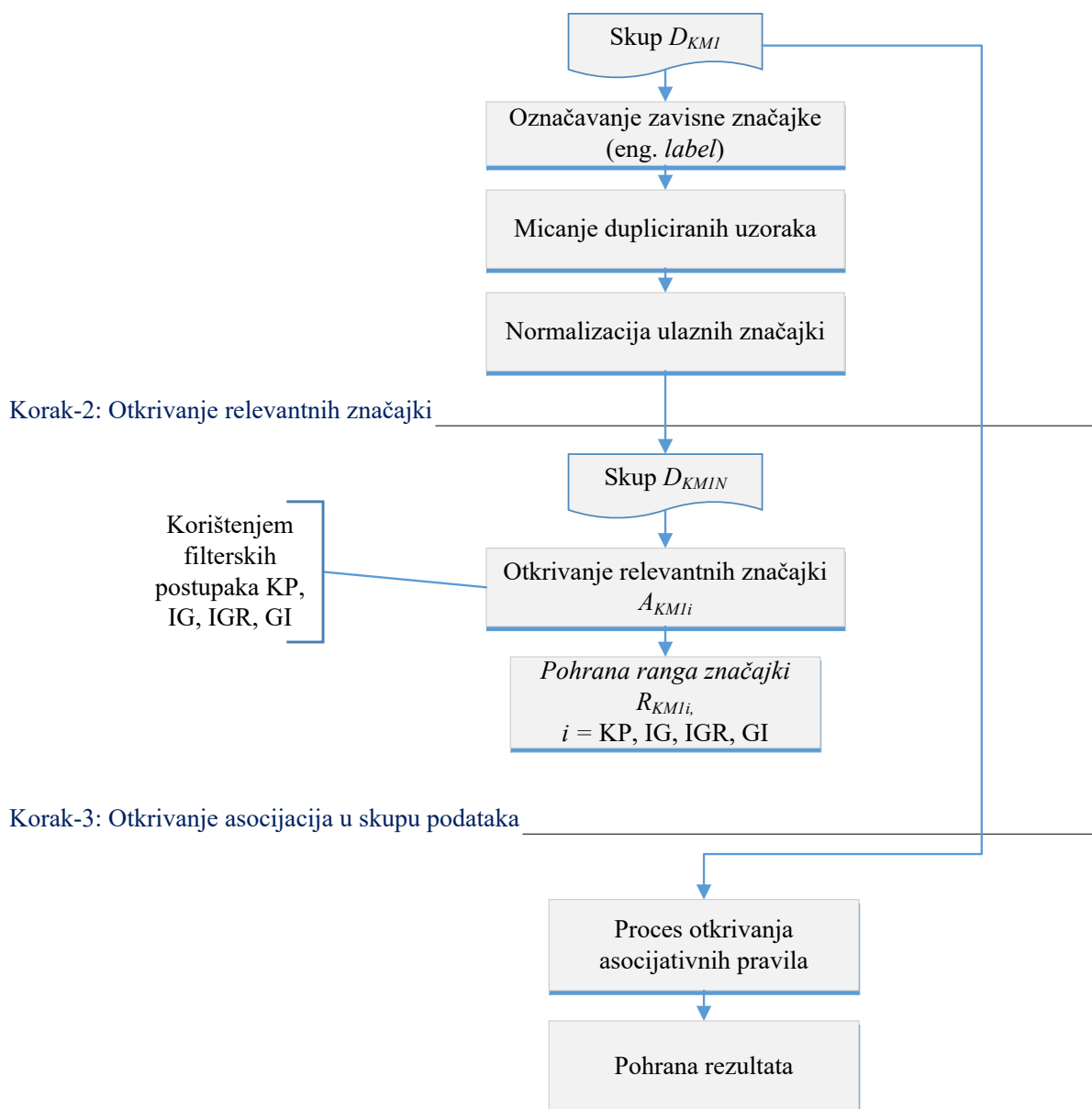
Proces za otkrivanje relevantnih ulaznih (nezavisnih) značajki s ciljem smanjenja dimenzionalnosti podataka proveden je u dva koraka za sve analizirane sustave  $S_i$  na razini flote zrakoplova, kako je prikazano na slici 4-1. Razlog objedinjavanja  $D_{KMI}$  skupova na razinu flote je taj što veći broj uzoraka rezultira točnijim rangom značajki i procjenom relevantnih značajki.

Glavni korak ovog procesa je opisan u odjeljku 4.1.1, gdje je prikazan okvir za primjenu filterskih postupaka korištenjem operatora ugrađenih unutar RapidMiner platforme.

Ulaz u proces je skup podataka  $D_{KMI}$ , koji se sastoji od niza uzoraka  $d_{KMI}$  i svaki uzorak je predstavljen kao  $n$ -dimenzionalni vektor  $X_{KMI} = (x_{KMI_1}, \dots, x_{KMI_n})$ , gdje je  $x_{KMI_i}$  do  $x_{KMI_{n-1}}$  vrijednost nezavisne značajke  $A_{KMI_i}$  za zadani uzorak, dok je  $x_{KMI_n}$  vrijednost zavisne značajke  $A_{KMI_n} = 1$  ili 0. Izlaz iz procesa je rang značajki  $R_{KMI_i}$ , gdje je  $i$  odabrani filterskih postupak.

Iako u domenskom problemu vrijedi pretpostavka o nezavisnosti ulaznih značajki, nad skupom podataka  $D_{KMI}$  dodatno je izgrađen model za otkrivanje asocijativnih pravila s ciljem pronalaženja značajki koje se pojavljuju zajedno.

Korak-1: Priprema skupa podataka



Slika 4-1. Dijagram toka procesa za otkrivanje relevantnih značajki i asocijacija unutar  $D_{KMIN}$  skupova podataka

#### 4.1.1. Priprema skupa podataka

Primjenom operatora *Read Excel*, koji omogućava povezivanje na vanjske izvore podataka u .xls(x) formatu, učitani su skup podataka  $D_{KMI}$ .

U učitanoj skupu podataka označena je (eng. *labeled*) zavisna značajka  $x_{KMI_n}$  (operator *SetRole*), nakon čega je uslijedila pretvorba tipa podatka zavisne značajke iz numeričkog u kategorijski tip (*true/false*) (operator *Numerical to Binominal*).

*Remove Duplicates* operator poslužio je za uklanjanje svih uzoraka iz skupa  $D_{KMI}$  koji imaju jednake vrijednosti vektora  $X_{KMI}$ .

Budući da vrijednosti uzorka unutar vektora  $X_{KMI}$  poprimaju različite raspone, provedena je normalizacija ulaznih značajki (operator *Normalize*) unutar skupa  $D_{KMI}$  (slika 4-2).

02/27-01	02/27-02	02/27-04	02/27-05	02/27-06	...	02/27-28	02/27-29	02/27-30	PI- zapis? [1 ili 0]
0	0	0	0,166666667	0	...	0	0	0	false
0	0,25	0,5	0,333333333	0	...	0	0	0	false
0	0	0	0,166666667	0	...	0	0	0	false
0	0,5	1	1	0,333333333	...	0	0	0	false
0,25	0	0	0	0	...	0,333333333	0	0	false
0,25	0	0	0	0	...	0,333333333	0	0	false
0	0	0	0	0	...	0	0	0,1	false
0	0	0	0	0	...	0	0	0	false
0	0	0	0	0	...	0,333333333	0	0	false
0	0	0	0	0	...	0,333333333	0	0,1	false
0	0	0	0	0	...	0,333333333	0	0	false
0	0	0	0	0	...	0	0	0	false
0	0	0	0	0	...	0	0	0	true
...	...	...	...	...	...	...	...	...	...

Slika 4-2. Prikaz skupa podataka  $D_{KMIN}$  za sustav  $S_i$

Ulazne značajke unutar skupa  $D_{KMI}$  skalirane su u jednaki raspon (od 0 do 1) i time je stvoren skup podataka  $D_{KMIN}$  koji će biti ulaz u korak-2.

#### 4.1.2. Otkrivanje relevantnih značajki

Sukladno zaključcima iz odjeljka 2.3.4., u ovome koraku će se koristiti četiri filterska postupka za otkrivanje relevantnih značajki; korelacijski postupak (KP), informacijski dobitak (IG), omjer informacijskog dobitka (IGR) i gini indeks (GI).

Proces koji opisuje cijeli korak-2 je implementiran unutar RapidMiner platforme i prikazan je u tablici 4-1.



Tablica 4-1. *Proces otkrivanja relevantnih značajki*

---

**Ulaz:** Pripremljeni skup podataka  $D_{KMIN}$ ,  $A_{KMINi}$  nezavisnih značajki

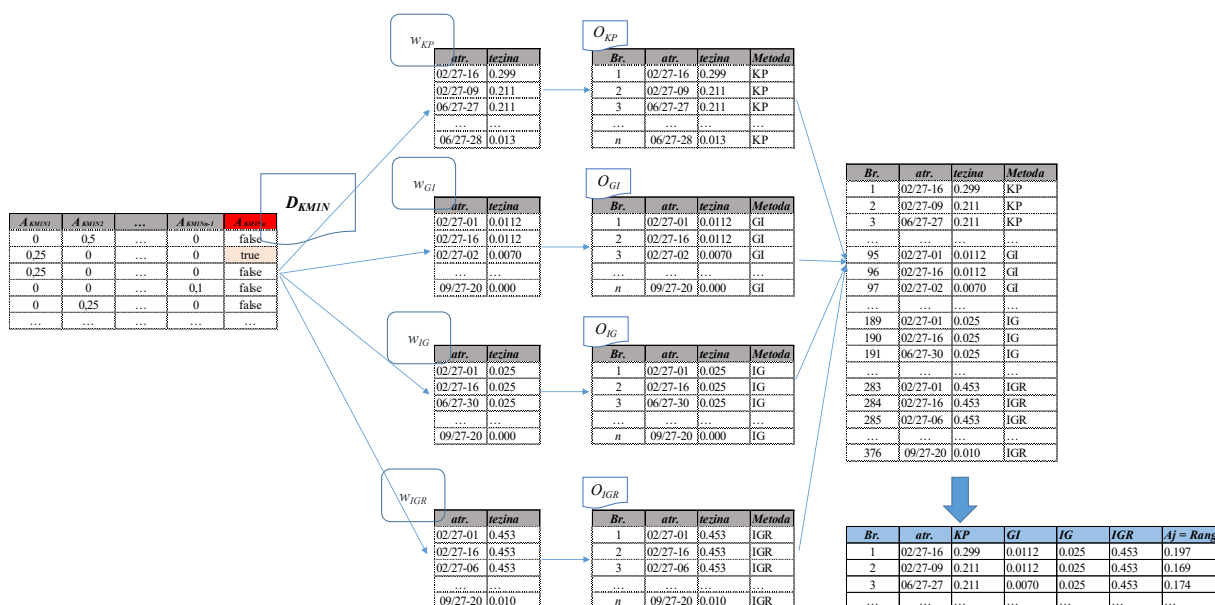
**Izlaz:** Rang značajki  $R_{KMi}$ ,  $i = KP, IG, IGR, GI$

---

Proces:

1. Za svaku  $A_{KMINi}$  nezavisnu značajku odredi težinu  $w_i$  temeljem postupaka KP, IG, IGR i GI // operator *Weight by Correlation, Weight by Gini Index, Weight by Information Gain, Weight by Information Gain Ratio*
  2. Zapamti težine  $w_i$  značajki za KP, IG, IGR i GI // operator *Remember*
  3. Pretvori težine  $w_i$  u izlazni set podataka  $O_i$  za svaki postupak// operator *Weights to Data*
  4. Za svaki izlazni set podataka  $O_i$  generiraj novi atribut  $A_{oi}$  i pridodaj mu naziv postupka // operator *Generate Attributes*
  5. Spoji sve izlazne setove podataka  $O_i$  // operator *Append*
  6. Transformiraj izlazne setove podataka  $O_i$  // operator *Pivot*
  7. Stvori novi atribut  $A_j$  koji će sadržavati srednju vrijednosti težina dobivenih temeljem svih filterskih postupaka za svaki  $A_{KMINi}$  // operator *Generate Aggregation*
  8. Normaliziraj sve vrijednosti unutar atributa  $A_j$  // operator *Normalize*
  9. Sortiraj vrijednosti težina prema padajućem nizu atributa  $A_j$  // operator *Sort*
  10. Presloži attribute abecednim redom// operator *Reorder Attributes*
- 

Za lakše razumijevanje procesa unutar tablice 4-1, na slici 4-3 dan je i grafički prikaz svih koraka unutar procesa.



Slika 4-3. Grafički prikaz procesa za otkrivanje ranga značajki

Iz tablice 4-1 i slike 4-3 je vidljivo da je, osim vektora ranga značajki za svaki od filterskih postupaka, stvoren i vektor značajki (spremljen unutar generiranog atributa  $A_j$ ) koji predstavlja srednju vrijednost težina svih filterskih postupaka. Generiranjem kumulativnog vektora ranga značajki moguće je u daljnjoj analizi usporediti odstupanja u rangiranju značajki svakog od odabranih filterskih postupaka u odnosu na kumulativni rang.

Nadalje, pohranjivanjem ranga značajki  $R_{KMIi}$ , za svaki od navedenih filterskih postupaka, prilikom izgradnje modela moguće je promotriti kako smanjenje broja ulaznih značajki utječe na odabrane mjere. Drugim riječima, potrebno je provesti analizu koliko se značajki može ukloniti iz izvornog vektora značajki, a da se postignu dobre vrijednosti statističkih mjera.

Preporuča se ispitivanje klasifikacijske mogućnosti skupova od 100%, 50%, 40%, 30%, 20%, 10% i 5% izvornog skupa značajki, no ovisno o broju značajki mogu se ispitati i neki drugi postotni udjeli [34].

Analizirajući vektore ranga značajki za KP, GI, IGR i GI postupke, u ovoj disertaciji ispitat će se klasifikacijske mogućnosti skupova od 20% do 80% izvornog skupa značajki. Razlog tome je što je ustanovljeno da posljednjih 20% značajki, rangiranih na temelju izvornog skupa, nema veliku težinsku vrijednost dobivenu odabranim filterskim postupcima (v. prilog D). Za domenski problem se smatra da je prvih 20% značajki relevantno, pa je prag za minimalni broj značajki u skupu postavljen na 20%.

#### 4.1.3. Otkrivanje asocijacija unutar značajki

Kao što je navedeno u odjeljku 2.3.3.2, asocijativna pravila ukazuju na učestalu vezu između vrijednosti dviju značajki u istom skupu. Iz tog razloga je unutar RapidMiner platforme razvijen proces za otkrivanje asocijativnih pravila nad skupom  $D_{KMI}$  (tablica 4-2).

Iz skupa  $D_{KMI}$  filtrirani su samo oni  $d_{KMI}$  uzorci gdje je zavisna značajka  $A_{KMI_n}$  poprimila vrijednost  $x_{KMI_n} = 1$ . Iako su značajke nezavisne, prilikom pojave određene značajke koja implicira nepravilnost u radu sustava, moguće je da se u kombinaciji s tom značajkom pojavi i neka druga značajka.

Ova pojava je povezana s primarnim i sekundarnim kvarovima, tj. zbog međusobne povezanosti komponenata unutar sustava, otkaz jedne komponente može uzrokovati i otkaz druge komponente.

Tablica 4-2. Proces otkrivanja asocijativnih pravila

<b>Ulaz:</b> Pripremljeni skup podataka $D_{KMI}$ , $A_{KMI_i}$ nezavisnih značajki
<b>Izlaz:</b> Asocijativna pravila
Postupak:
1. Iz skupa podataka $D_{KMI}$ odaberi $d_{KMI}$ uzorke gdje je vrijednost zavisne značajke $A_{KMI_n} = 1$ // operator <i>Filter Examples</i>
2. Pretvori vrijednosti svih značajki iz numeričkih u diskretne // operator <i>Numerical to Binominal</i>
3. Pronađi česte uzorke (eng. <i>frequent itemsets</i> ) unutar skupa podataka // operator <i>FP-Growth</i>
4. Stvori asocijativna pravila // operator <i>Create Association Rules</i>

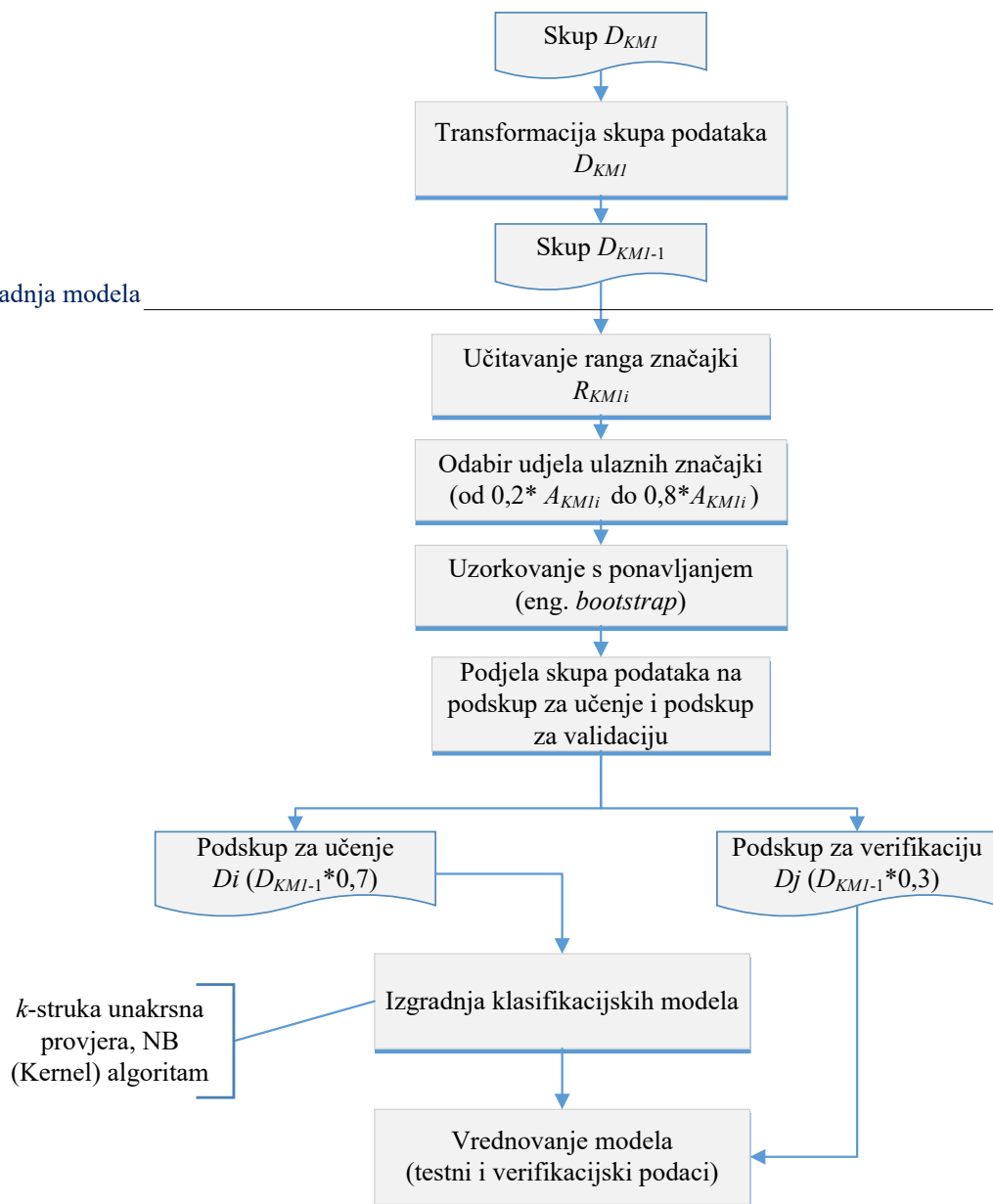
Prije otkrivanja čestih uzoraka unutar skupa podataka, operator *Numerical to Binominal* treba biti primijenjen kako bi se sve vrijednosti značajki pretvorile iz numeričkih u diskretne. Za otkrivanje asocijativnih pravila nije bitno koliku vrijednost ima određena značajka već da li se pojavila u uzorku. Ovdje je razlika između asocijativnih pravila i klasifikacijskih modela. Dok je vrijednost ulazne značajke presudna za odlučivanje u procesu klasifikacije, kod asocijativnih pravila je važno samo da je ta vrijednost istinita/neistinita (true/false). U domenskom slučaju to znači da će asocijativna pravila biti otkrivena među onim značajkama (porukama upozorenja po fazama leta) koje su u istom zapisu poprimile vrijednost „true.“

Za otkrivanje „jakih“ pravila potrebno je odrediti dva parametra – minimalnu podršku (eng. *minimum support*) i minimalnu pouzdanost (eng. *minimum confidence*). Oba parametra mogu poprimiti vrijednosti u rasponu od 0 do 1, a istraživači sami odlučuju prag za minimalne vrijednosti parametara. Budući da se u domenskom problemu žele otkriti sva asocijativna pravila unutar značajki, prag parametra minimalna podrška je postavljen na 0,1 (10%). Visoka vrijednost praga parametra minimalna podrška ne bi rezultirala otkrivanjem asocijativnih pravila iz razloga što nema veliki broj zapisa u skupu podataka za učenje. Vrijednost praga parametra minimalna pouzdanost je postavljena na 0,5 (50%).

To će rezultirati stvaranjem asocijativnih pravila onih značajki koje se u skupu pojavljuju zajedno u minimalno 10% zapisa i čija je kombinacija u minimalno 50% zapisa rezultirala da zavisna značajka poprimi vrijednost = 1.

## **4.2. Formulacija procesa za izgradnju klasifikacijskih modela prve grupe**

Izgradnja klasifikacijskih modela složeni je postupak koji nastupa nakon sastavljanja skupova podataka i izlučivanja značajki. Klasifikacijski modeli prve grupe će biti izgrađeni za svaki zrakoplov  $Ac_i$  na razini analiziranih sustava  $S_i$  i davati odgovor na pitanje „*Može li se očekivati da će u nadolazećem razdoblju nastati pilotski zapis u TLB temeljem prethodno nastale poruke upozorenja u određenoj fazi leta?*“ Slika 4-4 prikazuje proces za izgradnju klasifikacijskih modela prve grupe koji su provedeni unutar RapidMiner platforme u dva koraka.



Slika 4-4. Dijagram toka procesa izgradnje klasifikacijskih modela prve grupe

#### 4.2.1. Priprema skupa podataka

Poznato je da svaki  $d_{KMI}$  uzorak iz skupa  $D_{KMI}$  sadržava  $A_{KMIi}$  nezavisnih značajki čije vrijednosti  $x_{KMI1}$  do  $x_{KMI_{n-1}}$  mogu rezultirati da zavisna značajka  $A_{KMI_n}$  poprimi vrijednost  $x_{KMI_n} = 1$ . Stoga je potrebno transformirati skup podataka  $D_{KMI}$  na način da se prepoznaju vrijednosti nezavisnih značajki  $A_{KMIi}$  koje prethode tome da zavisna značajka  $A_{KMI_n}$  poprimi vrijednost  $x_{KMI_n} = 1$ . Drugim riječima, potrebno je prepoznati onu kombinaciju poruka upozorenja po fazama leta nastalih prije kombinacije poruka upozorenja po fazama leta koja je rezultirala

pilotskim zapisom u TLB. Proces za transformaciju skupa  $D_{KMI}$ , koji je razvijen unutar RapidMiner platforme, je prikazan u tablici 4-3.

Tablica 4-3. Prikaz procesa za stvaranje  $D_{KMI-1}$  skupa podataka

<b>Ulaz:</b> Skup podataka $D_{KMI}$
<b>Izlaz:</b> Skup podataka $D_{KMI-1}$
Proces:
1. Učitaj skup podataka $D_{KMI}$ // operator <i>Read Excel</i>
2. Filtriraj uzorke $d_{KMI}$ za svaki zrakoplov $Ac_i$ // operator <i>Filter Examples</i>
3. Dupliciraj izlazni skup // operator <i>Multiply</i>
4. Razdijeli skup $D_{KMI}$ na dva podskupa - podskup $D_{CI}$ gdje je zavisna značajka $A_{KMI_n}$ poprimila vrijednost $x_{KMI_n} = 1$ i podskup $D_{C0}$ gdje je zavisna značajka $A_{KMI_n}$ poprimila vrijednost $x_{KMI_n} = 0$ // operator <i>Filter Examples</i>
5. Grupiraj sve zapise po značajki <i>Datum</i> unutar podskupa $D_{C0}$ i stvori novi podskup $D_{C0g}$ // operator <i>Aggregate</i>
6. Spoji zapise iz podskupa $D_{C0g}$ i zapise iz podskupa $D_{CI}$ i stvori novi skup $D_{KMI-1}$ // operator <i>Append</i>
7. Eksportiraj skup $D_{KMI-1}$ u .xls datoteku // operator <i>Write Excel</i>

Izlazna datoteka, koja predstavlja novostvoreni skup  $D_{KMI-1}$  prikazana je na slici 4-5.

<i>Datum</i>	02/22-10	02/22-11	02/22-12	...	08/22-3	09/22-3	<i>PI-zapis? [1 ili 0]</i>	<i>Predviđa li se zapis na sljedećem letu?</i>
12.10.2009	0	0	0	...	0	0	0	0
18.10.2009	0	0	0	...	0	0	0	0
<b>01.11.2009</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>...</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>
18.11.2009	2	0	2	...	0	0	1	0
23.11.2009	0	0	0	...	0	0	0	0
04.12.2009	0	0	0	...	0	0	0	0
<b>06.12.2009</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>...</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>
07.12.2009	0	0	0	...	0	0	1	0
16.12.2009	0	0	0	...	0	0	0	0
....	....	....	....	....	....	....	....	....

Slika 4-5. Prikaz skupa podataka  $D_{KMI-1}$

Kao što je vidljivo iz tablice 4-3 i slike 4-5, prilikom stvaranja  $D_{KMI-1}$  skupa podataka grupirani su svi događaji (broj zabilježenih poruka upozorenja po fazama leta) koji su prethodili tome da zavisna značajka *Zapis? [1 ili 0]* poprimi vrijednost 1. Nakon grupiranja zapisa, stvorena je nova značajka *Predviđa li se zapis na sljedećem letu?*. Ta značajka služi za prepoznavanje uzoraka koji su prethodili zapisu u TLB, stoga će biti korištena kao zavisna značajka kod izgradnje modela prve grupe.

#### 4.2.2. Izgradnja modela

U ovome koraku izgradit će se modeli za svaki analizirani sustav  $S_i$  zrakoplova  $Ac_i$ .

Sukladno prethodno izvedenim zaključcima iz odjeljka 4.1.2, prilikom izgradnje modela provjerit će se kako smanjenje broja ulaznih značajki utječe na učinkovitost modela (odabrane mjere za vrednovanje klasifikacijskog modela).

Time će se iterativno smanjivati vektor ulaznih značajki  $A_{KMI}$  korištenjem operatora *Select by Weights*. U svakoj iteraciji će se filtrirati ulazne značajke u udjelu od 20% do 80% temeljem učitanog ranga značajki  $R_{KMI}$  (operator *Recall*), gdje je  $i = KP, IG, IGR$  i  $GI$ .

Operator *Bootstraping* će se koristiti za balansiranje podataka, tj. za početno uzorkovanje s ponavljanjem uzoraka slabe klase. Funkcija ovog operatora je da se smanji neuravnoteženost između klasa u analiziranom skupu podataka. Unutar ovog operatora moguće je odabrati dva načina uzorkovanja podataka – apsolutno uzorkovanje (eng. *absolute sampling*) i relativno uzorkovanje (eng. *relative sampling*). Apsolutno uzorkovanje podrazumijeva odabir podjednakog broja uzoraka iz obje klase (npr. 100 uzoraka iz slabe klase i 100 iz jake klase) te se na taj način stvara u potpunosti uravnotežen skup podataka za učenje modela. Relativno uzorkovanje podrazumijeva udio uzoraka koji treba biti odabran iz svake klase u danom skupu podataka, tj. ako se odabere 0,1 (10%), novonastali skup podataka će sadržavati 10% uzoraka iz slabe klase i 10% uzoraka iz jake klase. Za izradu modela u ovoj disertaciji će se koristiti relativno uzorkovanje gdje će se odabrati uzorkovanje primjeraka iz obje klase u omjeru od 100% (1,0). Izlaz iz ovog operatora će biti skup podataka koji će imati isti broj uzoraka kao i izvorni skup podataka, no neki uzorci iz slabe klase će biti odabrani više puta. Omjer klasa u tom skupu i dalje neće biti podjednak, tj. jaka klasa će i dalje biti većinska kako bi skup i dalje predstavljao realan problem.

Nakon toga, primjenom operatora *Split* skup podataka će se podijeliti na dva podskupa – podskup za učenje modela  $Di$  i podskup za verifikaciju  $Dj$ . Podjela na podskupe  $Di$  i  $Dj$  provest će se s različitim vrstama uzorkovanja podataka – stratificiranim uzorkovanjem i slučajnim izborom (v. odjeljak 2.3.5.1 za objašnjenje).

Temeljem zaključaka iz odjeljka 3.5, prošireni NB klasifikacijski algoritam (NB (kernel)) će biti primijenjen za izgradnju modela, a za procjenu pogreške modela koristit će se 10-struka unakrsna validacija (operator *Cross Validation*).

Kao što je prikazano u odjeljku 3.5, za nominalne/kategoričke vrijednosti značajki izvode se vjerojatnosti za kombinaciju vrijednosti značajki jednostavnim prebrojavanjem mogućih vrijednosti i dijeljenjem tih vrijednosti s brojem svih mogućnosti. No, stvari se mijenjaju kada značajka poprima numeričke vrijednosti, što je slučaj u domenskom problemu. Kod jednostavne implementacije, vjerojatnosti za numeričke vrijednosti su izvedene iz jedne distribucije (obično Gaussove) koja se primjenjuje na podatke.

Distribucija bazirana na procjeni gustoće vjerojatnosti (eng. *kernel-based distribution*) zamjenjuje jednostavnu jedno-modalnu distribuciju s distribucijom koja se sastoji od dodatnog sloja (eng. *additive overlay*) višestrukih Gaussovih distribucija (v. sliku 2-8).

NB (kernel) algoritam, ugrađen unutar RapidMiner alata, ima dva parametra koji procjenjuju kako će se model ponašati, tj. kako je procjena funkcije gustoće vjerojatnosti odrađena. Tako je potrebno specificirati minimalnu propusnost (širinu) jezgre (eng. *minimum bandwidth*) i broj jezgara, a zadane vrijednosti su 0,1 (minimalna propusnost) i 10 (broj jezgara) [98].

Broj jezgara je jednostavno broj distribucija koji će se koristiti. Ukoliko je taj broj prevelik, distribucija postaje kompleksnija i može rezultirati generalizacijom modela nad skupom podataka za učenje (eng. *overfitting*). Ukoliko je broj premalen, postoji mogućnost da se zanemare vršne vrijednosti distribucije (eng. *data peaks*).

Parametar minimalna propusnost jezgre predstavlja širinu svake jezgre. Ukoliko je jezgra preširoka, to će rezultirati uglađenijom (eng. *smoother*) krivuljom distribucije.

Stoga je potrebno odrediti optimalne vrijednosti ova dva parametra. Optimizacija vrijednosti će se provesti primjenom operatora *Optimize Parameters (Grid)*. Budući da u literaturi nisu pronađene preporučene vrijednosti, a ni rasponi za ova dva parametra, temeljem informacija dobivenih od strane proizvođača<sup>4</sup> RapidMiner alata, raspon za broj jezgara je postavljen od 1 do 10. Također, prema preporukama proizvođača, kod određivanja minimalne propusnosti jezgre, odabrano je da vrijednosti budu između 0,1 i 0,5. Primjenom ovog operatora, za svaki skup podataka, neovisno o udjelu ulaznih značajki, će biti odabrane optimalne vrijednosti parametara.

Naposljetku, primjenom operatora *Apply Model*, validacijski podaci će biti primijenjeni na već izgrađeni model kako bi se testirale prediktivne sposobnosti modela. Operator *Performance* će

---

<sup>4</sup> Kontakt ostvaren elektroničkim putem



zabilježiti sve vrijednosti statističkih mjera koje će služiti za evaluaciju modela. Osim vrijednosti statističkih mjera, izlaz iz operatora *Performance* je i zavisna značajka koja klasificira ulazni skup značajki u pozitivnu ili negativnu klasu. Za svaku klasu su izračunate i „sigurnosti“ predviđanja (eng. *prediction confidence*), tj. vjerojatnosti da će ulazni primjerci biti svrstani u predviđenu klasu (slika 4-6).

Row No.	Predviđa li se zapis na sljedećem letu?	prediction (Predviđa li se zapis na sljedećem letu?)	confidence (false)	confidence (true)	02/22-10	02/22-11	02/22-12	...	08/22-3	09/22-3
1	TRUE	TRUE	0,300	0,700	0	0	0	...	0	0
2	FALSE	FALSE	0,759	0,241	2	0	2	...	0	0
3	FALSE	TRUE	0	1	0	0	0	...	0	0
4	FALSE	FALSE	1	0	0	0	0	...	0	0
5	TRUE	TRUE	0,300	0,700	0	0	0	...	0	0
6	FALSE	FALSE	1	0	0	0	0	...	0	0
7	FALSE	FALSE	1	0	0	0	0	...	0	0
....	....	....	....	....	....	....	....	....	....	....

Slika 4-6. Prikaz izlaznih rezultata iz klasifikacijskog modela prve grupe

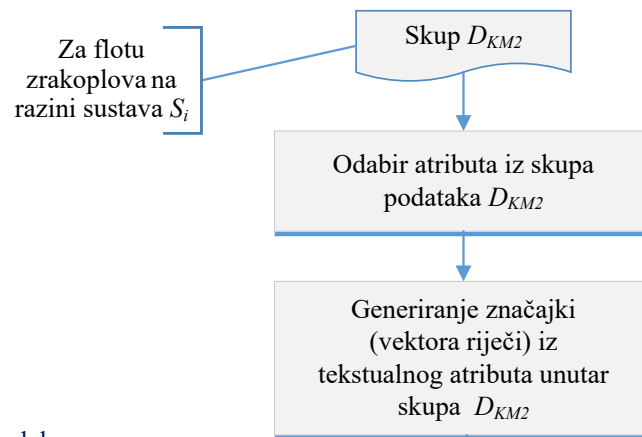
Ovakav prikaz izlaza može poslužiti donositelju odluke kao potpora pri odlučivanju kod planiranja prijevremenih aktivnosti održavanja (testiranje ili popravak komponente) temeljem poruka upozorenja koje je određeni sustav generirao.

### 4.3. Formulacija procesa za izgradnju klasifikacijskih modela druge grupe

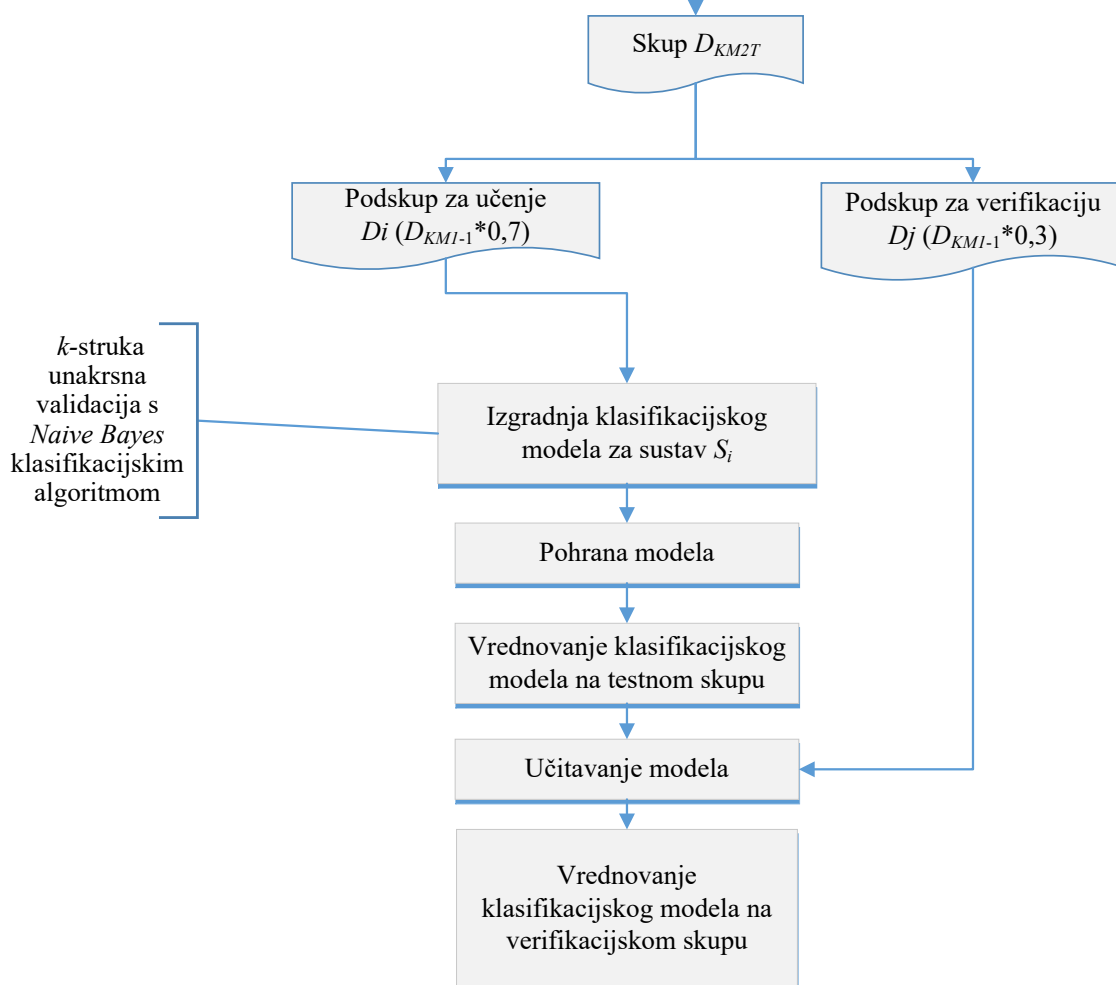
Izgradnja klasifikacijskih modela druge grupe provedena je unutar RapidMiner platforme i obuhvaća klasifikaciju tekstualnih zapisa na skupu podataka  $D_{KM2}$  (slika 3-13). Model se sastoji od dva koraka (slika 4-7), koji su provedeni usvajanjem metodologije iz rada [16]. Za razliku od modela u radu [16], koji koristi samo jedan izvor podataka i klasifikacijski algoritam neuronske mreže (NN), u ovoj disertaciji će se koristiti udruženi skup podataka  $D_{KM2}$  i klasifikacijski algoritam Naivni Bayes (NB). Klasifikacijska točnost modela u ovoj disertaciji će se usporediti s klasifikacijskom točnošću modela iz rada [16], što će biti prikazano u poglavlju 5.

Modeli druge grupe će biti izgrađeni za flotu zrakoplova na razini analiziranih sustava  $S_i$  i davat će odgovor na pitanje: „Može li se očekivati da će zabilježeni tekstualni pilotski zapis utjecati na raspoloživost zrakoplova?“

Korak-1: Priprema skupa podataka



Korak-2: Izgradnja modela



Slika 4-7. Dijagram toka procesa izgradnje klasifikacijskih modela druge grupe

#### 4.3.1. Priprema skupa podataka

Iz skupa podataka  $D_{KM2}$  koji se sastoji od značajki  $A_{KM2} = \{A/C, Date, ATA, Description, Utjecaj na raspoloživost, Posljedica\}$ , primjenom operatora *Select Attributes*, odabrana je nezavisna značajka *Description* koja sadržava tekstualne zapise o prošlim kvarovima ili nepravilnostima u radu sustava i operacijskim zastoјima, dok su značajke *Utjecaj na raspoloživost* i *Posljedica* odabrane kao zavisne značajke.

Odabirom operatora *Process Documents from Data*, unutar kojeg su inkorporirani operatori *Tokenize*, *Transform Cases*, *Filter Stopwords (English)* i *Filter Tokens (by Length)*, nestrukturirani tekstualni zapisi pretvoreni su u vektore riječi na temelju kojih je stvoren skup podataka  $D_{KM2T}$ .

Tekstualni zapisi, koji se učitavaju iz datoteke mogu se opisati kao matrica  $Z$  veličine  $N \times p$ , gdje je  $N$  broj dokumenata, a  $p$  je broj ključnih riječi (pojmovi) iz unije svih dokumenata. Pojmovi su definirani kao riječi koje su informativne i pružaju znanja o sadržaju unutar dokumenta. Kako bi se iz matrice maknuli pojmovi koji nisu značajni za analizu, obično se primjenjuje redukcijska mjera pod nazivom TF-IDF (eng. *Term Frequency Inverse Document Frequency*) [99].

TF-IDF mjera se izračunava pomoću frekvencije pojmova u dokumentu TF i inverzne frekvencije pojmova u skupu dokumenata IDF (provjerava broj dokumenata u kojima se pojavio konkretan pojam). Najčešće upotrebljavane formule za izračunavanje TF-IDF mjere su [100]:

$$tf-idf(p) = tf(p) \cdot idf(p) \quad (4.1)$$

$$idf(p) = \log \left( \frac{N}{df(p)} \right), \quad (4.2)$$

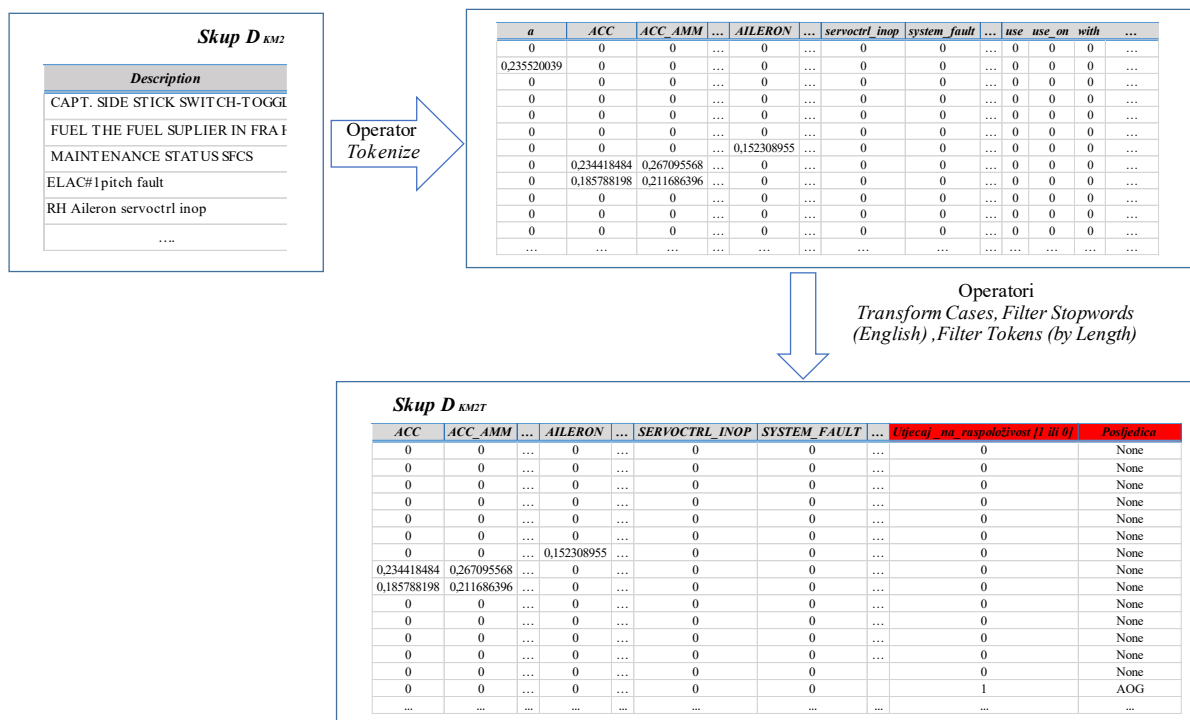
gdje  $p$  predstavlja pojam,  $tf(p)$  prikazuje broj pojavljivanja tog pojma u dokumentu,  $df(p)$  je broj dokumenata u skupu dokumenata koji sadrže taj pojam, a  $N$  je broj dokumenata u skupu. Ukoliko se pojam pojavi u velikom broju dokumenata, smatra se nevažnim i pridružena težina je niža (blizu nuli). Kada je pojam jedinstven i pojavljuje se u nekoliko dokumenata, pridružena težina se kreće bliže jedinici. Ova metrika je primjerena u slučaju kada se ne zahtijeva dodatna lingvistička analiza teksta [100].

TF-IDF je mjera za čitav niz dokumenata, dok je TF mjera (još nazvana lokalna težina pojmova) mjera za jedan dokument [101].

Osim TF-IDF mjere, unutar operatora *Process Documents from Data*, moguće je odabrati i samo TF mjeru, TO (eng. *Term Occurences*) i BTO (eng. *Binary Term Occurences*) mjeru za stvaranje vektora pojmove. TO mjera predstavlja apsolutni broj ponavljanja pojma u dokumentu, dok BTO mjera rezultira binarnim vrijednostima za određeni pojam (ako se pojam pojavio =1, ako se pojam nije pojavio = 0).

Budući da TD-IDF mjera eliminira pojmove koji se često ponavljaju, a u nekim slučajevima su bitni za domenski problem, ta mjera neće bi korištena u radu. Umjesto te mjere koristit će se TF mjera, dok će zadatak prepoznavanja važnih/nevažnih pojmove biti prepušten klasifikacijskom algoritmu u koraku izgradnje modela.

Grafičko objašnjenje procesa pretvaranja nestrukturiranih zapisa u skup podataka za učenje klasifikacijskog modela je prikazano slikom 4-8.



Slika 4-8. Prikaz stvaranja skupa  $D_{KM2T}$

Iz slike 4-8 mogu se razlučiti funkcije operatora *Tokenize*, *Transform Cases*, *Filter Stopwords (English)* i *Filter Tokens (by Length)*.

Primjenom operatora *Tokenize* provedena je tokenizacija, tj. proces raščlanjivanja teksta u izraze, riječi, simbole ili druge značajne elemente nazvane tokenima. Cilj tokenizacije je istražiti riječi (pojmove) u određenoj rečenici [102].

*Transform Cases* operator pretvara znakove (slova) svih riječi (tokena) iz malih u velika. Razlog tome je što je ista riječ unutar nestrukturiranih tekstualnih zapisa u nekim slučajevima (zapisima) zabilježena velikim slovima, a u nekim slučajevima malim slovima. Ovaj operator je također ključan kako bi se izbjeglo grupiranje istih pojmova koji trebaju biti pretvoreni u značajke. Drugim riječima, ukoliko se ne primijeni ovaj operator, za istu riječ, samo zapisanu malim i velikim slovima, bi se smatralo da predstavlja dva različita pojma (značajke).

Operator *Filter Stopwords (English)* služi za micanje pojmova koji predstavljaju članove riječi (primjer: *a, the, an*) i veznika (primjer: *and, or, nor, where, ...*). Svrha ovog operatora je da makne sve pojmove koji ne doprinose otkrivanju znanja.

Slično tome operatoru, operator *Filter Tokens (by Length)* miče sve pojmove koji imaju minimalni broj znakova 2, a maksimalni broj 25. Vrijednosti minimalnih i maksimalnih brojeva znakova su postavljene proizvoljno temeljem ručne analize teksta. Smatra se da pojmovi kraći od 2 znaka i dulji od 25 ne sadrže korisne informacije.

#### 4.3.2. Izgradnja modela

Stvaranjem skupa  $D_{KM2T}$  u istom procesu unutar RapidMiner platforme kreće izgradnja klasifikacijskog modela druge grupe. Kao što je vidljivo sa slike 4-8, ulazne značajke  $D_{KM2T}$  skupa predstavljaju pojmove koji se javljaju unutar tekstualnih zapisa i poprimaju vrijednosti dobivene odabranom mjerom za stvaranje vektora pojmova (TF mjera).

Skup  $D_{KM2T}$  će biti podijeljen na skup za izgradnju modela (testni skup podataka, 70% uzoraka) i skup za provjeru modela (verifikacijski skup podataka, 30% uzoraka).

Za razliku od skupa  $D_{KMI-1}$ ,  $D_{KM2T}$  skup sadržava dvije zavisne značajke; *Utjecaj na raspoloživost* i *Posljedica*. Zavisna značajka *Utjecaj na raspoloživost* je binarna (1/0), dok je značajka *Posljedica* nominalna (eng. *polynominal*) i sadržava klase „otkaz leta“ (eng. *cancellation*), „zastoj“ (eng. *delay*), „prizemljenje zrakoplova“ (eng. *aircraft on ground, AOG*) i „bez posljedice na raspoloživost zrakoplova“ (eng. *none*).

Budući da se u većini  $D_{KM2T}$  skupova za analizirane sustave nalazi manje od 100 uzoraka, skupovi će biti balansirani operatorom *bootstrapping* za slučaj kada će zavisna značajka biti *Posljedica*. Uzorkovanje primjeraka će biti provedeno relativnim uzorkovanjem, gdje će odabrati uzorkovanje primjeraka iz obje klase u omjeru od 200% (2,0), kako bi skupovi podataka sadržavali dovoljan broj uzoraka za učenje i verifikaciju modela.

Za svaku zavisnu značajku (labelu), model će biti izgrađen primjenom NB algoritma (operator *Naive Bayes*), a procjena pogreške modela će biti dobivena 10-strukom unakrsnom validacijom (operator *Cross Validation*).

Kao i kod izgradnje modela prve grupe, primjenom operatora *Apply Model*, validacijski podaci će biti primijenjeni na već izgrađeni model kako bi se testirale prediktivne sposobnosti modela. Operator *Performance* će zabilježiti sve vrijednosti statističkih mjera koje će služiti za evaluaciju modela i „sigurnosti“ predviđanja (eng. *prediction confidence*), tj. vjerojatnosti da će ulazni primjerci biti svrstani u predviđenu klasu.

#### **4.4. Zaključak poglavlja i smjernice za nastavak istraživanja**

Na temelju razrađenih procesa u prethodno prikazanim odjeljcima, zaključak ovog poglavlja je sveden na sljedeće točke:

1. Proces za otkrivanje relevantnih ulaznih značajki provest će se za analizirane sustave  $S_i$  na razini flote zrakoplova s ciljem postizanja veće točnosti kod dodjeljivanja težina svakoj značajki primjenom filterskih postupaka na objedinjenom skupu podataka.
2. Evaluacija ranga značajki dobivenih filterskim postupcima će biti provedena prilikom izgradnje klasifikacijskih modela prve grupe na način da će se iz izvornog vektora značajki smanjivati udio ulaznih značajki i promatrati evaluacijske mjere novoizgrađenih modela.
3. Klasifikacijski modeli prve grupe će biti izgrađeni za svaki zrakoplov  $Ac_i$  na razini analiziranih sustava  $S_i$  kako bi se mogla predvidjeti pojava zapisa u TLB na razini pojedinog zrakoplova i sustava u nadolazećem vremenskom periodu.
4. Temeljem zaključaka iz odjeljka 3.5 i 4.2.2, kod izgradnje klasifikacijskih modela prve grupe provest će se balansiranje podataka s relativnim uzorkovanjem gdje će se odabrati uzorkovanje primjeraka iz obje klase u omjeru od 100% (1,0). Izlaz će biti skup podataka koji će imati isti broj uzoraka kao i izvorni skup podataka, no dio uzoraka iz slabe klase će biti odabran više puta.
5. Analogno procesu otkrivanja relevantnih ulaznih značajki, klasifikacijski modeli druge grupe će biti izgrađeni za flotu zrakoplova na razini analiziranih sustava  $S_i$  s ciljem točnije klasifikacije tekstualnih zapisa na skupu podataka s većim brojem uzoraka.
6. Budući da se u većini skupova za izgradnju klasifikacijskih modela druge grupe nalazi manje od 100 uzoraka, provest će se relativno uzorkovanje primjeraka u omjeru od 200% (2,0), kako bi skupovi podataka sadržavali dovoljan broj uzoraka za učenje i verifikaciju modela druge grupe

## 5. PRIKAZ I EVALUACIJA REZULTATA ISTRAŽIVANJA

U ovome poglavlju prikazani su rezultati procesa strukturiranih u poglavlju 4. U prvom odjeljku su prikazane odabrane mjere za evaluaciju rezultata klasifikacijskih modela. Radi boljeg razumijevanja ulaznih značajki, koje se koriste za izgradnju klasifikacijskih modela prve grupe (poruke upozorenja po fazama leta) i klasifikacijskih modela druge grupe (tekstualni opisi nepravilnosti u radu sustava ili kvarova), u drugom odjeljku su objašnjene glavne funkcije analiziranih sustava. Ovaj odjeljak također prikazuje i rezultate procesa za otkrivanje relevantnih značajki i asocijativnih pravila. U trećem i četvrtom odjeljku predstavljeni su rezultati dobiveni klasifikacijskim modelima na temelju čega je dan zaključak poglavlja u petom odjeljku.

### 5.1. Uvodna razmatranja

*Evaluacija rezultata filterskih postupaka.* Najzastupljeniji pristup evaluacije filterskih postupaka je promatranjem rezultata statističkih mjera na način da se smanjuje vektor ulaznih značajki, dobiven rangom odabranog filterskog postupka, na izgrađenom klasifikacijskom modelu [103]. Ukoliko je filterski postupak dobro rangirao relevantne značajke, klasifikacijski model sa smanjenim udjelom značajki bi trebao rezultirati boljim vrijednostima statističkih mjera od modela izgrađenog s ukupnim brojem ulaznih značajki. Za slučaj kada su relevantne značajke već poznate, dovoljna je stručna ekspertiza za potvrđivanje ranga značajki [103]. Temeljem procesa prikazanog u odjeljku 4.1.2 dobiven je rang značajki koji je prvo evaluiran od strane stručnjaka za održavanje. Ukupni rang značajki je stvoren za svaki analizirani sustav (ATA 22, ATA 27, ATA 29, ATA 32 i ATA 34) na temelju srednje težine svake značajke dobivene različitim filterskim postupcima te je potvrđena vjerodostojnost rezultata od strane stručnjaka za održavanje zrakoplova. Zbog manjih odstupanja kod dodjeljivanja težina pojedinim značajkama temeljem svakog filterskog postupka, dodatno je provedena evaluacija postupaka promatranjem rezultata statističkih mjera prilikom smanjenja udjela ulaznih značajki kod izgradnje klasifikacijskih modela prve grupe.

*Evaluacija klasifikacijskih modela prve grupe.* Na temelju zaključaka izvedenih iz odjeljka 3.5, zbog neuravnoteženosti skupova podataka, klasifikacijski modeli prve grupe će biti evaluirani pomoću tri mjere; *F*-mjere, *osjetljivosti* i *specifičnosti*. Iako u domenskoj literaturi nisu pronađene minimalne vrijednosti (pragovi) navedenih statističkih mjera, iz izraza za svaku mjeru prikazanih u odjeljku 2.3.5.2, može se zaključiti da model ima bolje prediktivne

sposobnosti ukoliko su vrijednosti navedenih mjera bliže 1,00 (100%). Kako bi se postavili pragovi svih triju mjera, prvo su izgrađeni klasifikacijski modeli na razini svakog sustava i zrakoplova s ukupnim brojem ulaznih značajki iz skupa za učenje  $D_{KMI-1}$ . Izgradnjom modela s ukupnim brojem ulaznih značajki uz primjenu različitih načina uzorkovanja podataka (uravnoteženog (eng. *stratified*, STR) ili slučajnog (eng. *shuffled*, SHU)), analizirane su vrijednosti statističkih mjera verifikacijskih skupova podataka. Model koji je rezultirao boljim vrijednostima statističkih mjera je odabran kao reprezentativan i te vrijednosti su postavljene kao minimalni pragovi svake mjere. S ciljem izgradnje modela koji će biti brži i bolje predviđati nove primjerke, ulazni broj značajki je iterativno smanjivan (od 20% do 80%). U svakoj iteraciji su zabilježene vrijednosti *F-mjere*, *osjetljivosti* i *specifičnosti* ovisno o načinu uzorkovanja podataka. Ukoliko je model sa smanjenim brojem ulaznih značajki rezultirao boljim vrijednostima statističkih mjera, tj. prešao minimalni prag svake mjere dobiven modelom izgrađenim s ukupnim brojem značajki, smatra se da taj model ima bolje generalizacijske sposobnosti za klasifikaciju novih primjeraka za dani skup podataka.

*Evaluacija klasifikacijskih modela druge grupe.* Mjera za evaluaciju klasifikacijskih modela druge grupe je *točnost* modela. Mjera točnosti je odabrana iz razloga kako bi se izgrađeni klasifikacijski modeli u ovoj disertaciji mogli usporediti s modelom izgrađenim u istraživanju [16], gdje je *točnost* odabrana kao glavna evaluacijska mjera modela i iznosila je približno 93%. Osim točnosti modela, za klasifikacijske modele druge grupe, prikazane su i matrice zabune (v. odjeljak 2.3.5.2) te su za svaku klasu (vrijednost koju zavisna značajka može poprimiti) izračunate i mjere *osjetljivosti (odziva)* i *preciznosti*.

## **5.2. Rezultati ranga relevantnih značajki i asocijativnih pravila**

Kao što je prikazano u odjeljku 4.1, ulaz u proces za otkrivanje ranga relevantnih značajki i asocijativnih pravila je skup podataka  $D_{KMIN}$ . Ulazne značajke unutar  $D_{KMIN}$  skupa podataka predstavljaju kombinaciju poruka upozorenja po fazama leta, a zavisna značajka

### *5.2.1. Rezultati ranga relevantnih značajki*

S ciljem razumijevanja poruka upozorenja koje se javljaju na razini analiziranih zrakoplovnih sustava, opisane su funkcije svakog sustava. U prilogu D nalaze se izvorno dodijeljene težine svim značajkama na razini analiziranih sustava. Na grafovima u nadolazećim odjeljcima prikazane su normalizirane težine radi lakše usporedbe rangova značajki dobivenih različitim



filterskim postupcima. Zbog bolje preglednosti ranga značajki, svi grafovi sadrže samo prvih 50% relevantnih značajki.

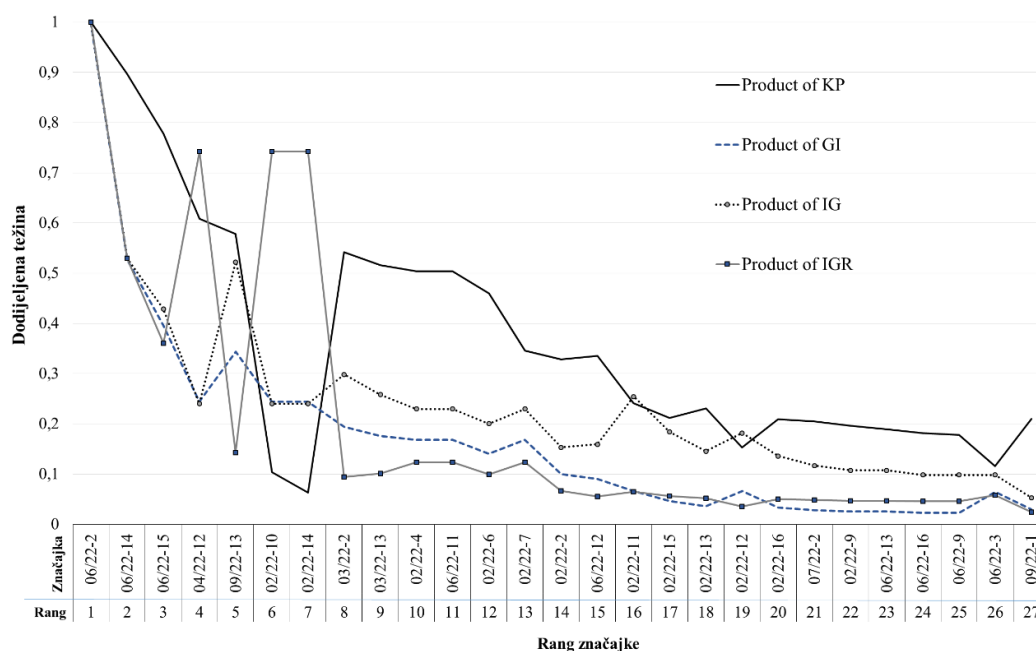
#### 5.2.1.1. ATA 22

Sustav automatskog leta (ATA 22, eng. *auto flight*) je važan sustav zrakoplova koji se sastoji od jedinica i komponenata koje kontroliraju smjer, pravac, visinu i brzinu leta zrakoplova. Sastoji se od dva glavna računala; FMGC računala (eng. *flight management i guidance comuter*) i FAC računala (eng. *flight augmentation computer*) [104].

FAC osigurava funkcije za upravljanje zrakoplovom po vertikalnoj osi; sustav stabilizacije skretanja (eng. *yaw damper*), ograničenje odklona kormila pravca (eng. *rudder travel limiting*), kompenzator kormila pravca (eng. *rudder trim*), autopilot skretanja (eng. *yaw autopilot*), itd. Ulazne informacije u FAC računalo se prikupljaju iz drugih sustava i komponenata kao što su podvozje i zakrilca/predkrilca zrakoplova, kako bi se uzele u obzir promjene u konfiguraciji zrakoplova, posebice tijekom polijetanja i slijetanja.

Glavna svrha FMGC računala je da osigura funkcioniranje autopilot sustava, funkciju (eng. *mode*) za plovību i prizemljenje, bočnu navigaciju (eng. *lateral navigation*), vertikalnu navigaciju (eng. *vertical navigation*) i praćenje performansi zrakoplova. Funkcije FMGC-a najčešće su upravljane pomoću MCDU-a (eng. *Multipurpose Control and Display Units*) i FCU-a (eng. *Flight Control Unit*). FMGC dobiva potrebne ulazne vrijednosti iz senzora i iz MCDU-a (npr. visinu leta (eng. *flight level*), brzinu leta, itd.). Uz senzorske podatke i plan leta koji je unio pilot putem MCDU-a, FMGC je u stanju izračunati pretpostavke za visinu, vrijeme i gorivo svake točke leta. FCU se koristi za odabir i kontrolu različitih načina letenja, aktiviranje autopilota (eng. *autopilot*, AP), jedinice za usmjeravanje zrakoplova po zadanoj putanji (eng. *flight director*, FD) i jedinice kontrole potiska motora (eng. *autothrust*, A/THR) [104].

Kombinacija svih poruka upozorenja po fazama leta (promatrane značajke) koje je generirao sustav ATA 22 u promatranom razdoblju nalazi se u prilogu D, a na slici 5-1 je prikazan rang za prvih 50% značajki. Značajke su rangirane prema srednjoj vrijednosti svih težina, dobivene nakon izračuna pojedinačnih težina temeljem KP, GI, IG i IGR filterskog postupka (v. odjeljak 4.1.2), za  $D_{KMIN}$  skup podataka ATA 22 sustava.



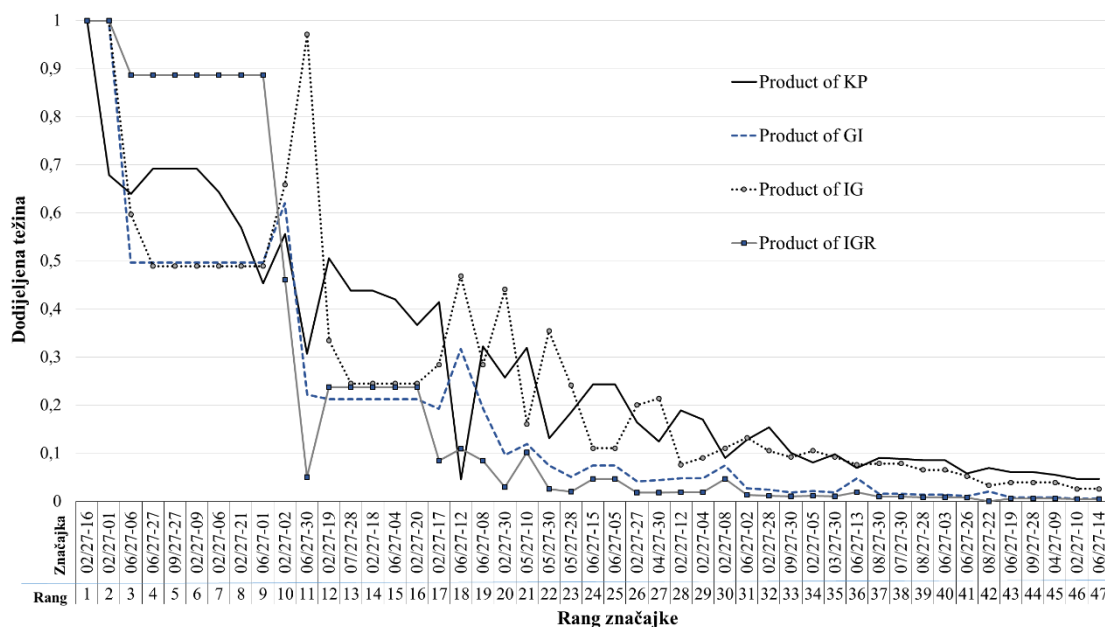
Slika 5-1. ATA 22 – prikaz ranga prvih 50% značajki

Iz slike 5-1 vidljivo je kako postoje manja odstupanja kod dodjeljivanja težina značajkama primjenom IGR i KP filterskih postupaka. Ta odstupanja su vidljiva kod rangiranja prvih 8 značajki, a nakon toga je uočen pravilan trend rangiranja značajki.

#### 5.2.1.2. ATA 27

Sustav kontrole leta (ATA 27, eng. *flight controls*) je podijeljen na primarne i sekundarne kontrole leta. Primarne kontrole leta (kontrolne površine) su potrebne za sigurno upravljanje zrakoplovom tijekom leta i sastoje se od krilca zrakoplova (eng. *aileron*), kormila visine (eng. *elevators*) i kormila pravca (eng. *rudder*). Krilca su smještena na izlaznim rubovima krila, elevator je smješten na horizontalnom stabilizatoru, a kormilo pravca je smješteno na vertikalnom stabilizatoru. Krilca zrakoplova kontroliraju gibanje oko uzdužne osi (valjanje, eng. *roll*), kormila visine rotaciju oko poprečne osi (propinjanje, eng. *pitch*) i kormilo pravca gibanje oko okomite osi (skretanje, eng. *yaw*). Sekundarne kontrole leta namijenjene su poboljšanju karakteristika performansi zrakoplova, a sastoje se od zakrilca zrakoplova (eng. *flaps*) i predkrilca (eng. *slats*), spojlera (eng. *spoilers*) i kompenzatora (eng. *trim systems*). Računala za kontrole leta primaju ulazne naredbe iz upravljačke palice (eng. *sidestick*), analiziraju ih i provjeravaju da se radi o sigurnim naredbama te koriste te naredbe za upravljanje kontrolnim površinama. Ta računala su dva ELAC računala (eng. *Elevator and Aileron Computers*), tri SEC računala (eng. *Spoiler and Elevator Computers*) i dva FAC računala (eng. *Flight Augmentation Computer*) [104].

Kombinacija svih poruka upozorenja po fazama leta koje je generirao sustav ATA 27 u promatranom razdoblju nalazi se u prilogu D, a na slici 5-2 je prikazan rang za prvih 50% značajki. Značajke su rangirane prema srednjoj vrijednosti svih težina, dobivene nakon izračuna pojedinačnih težina temeljem KP, GI, IG i IGR filterskog postupka (v. odjeljak 4.1.2), za  $D_{KMIN}$  skup podataka ATA 27 sustava.



Slika 5-2. ATA 27 – prikaz ranga prvih 50% značajki

Iz slike 5-2 vidljiv je pravilan trend dodjeljivanja težina značajkama do ranga 10 i iza ranga 30. Manja odstupanja kod dodjeljivanja težina značajkama su vidljiva kod značajki kojima se rang nalazi između ovog raspona.

### 5.2.1.3. ATA 29

Hidraulički sustav (ATA 29, eng. *hydraulic power*) koristi hidraulički fluid pod pritiskom za pogon mehaničkih komponenata. Hidraulički se sustav sastoji od hidrauličkog fluida plus tri glavne mehaničke komponente: generator tlaka (eng. *pressure generator*) ili hidraulička pumpa, hidraulički aktuator koji pogoni komponente i sustav cjevovoda (eng. *plumbing system*) koji sadržava i po potrebi distribuira (eng. *channels*) hidraulički fluid kroz zrakoplov. Ovisno o tipu zrakoplova, postoji minimalno jedan, odnosno dva ili tri hidraulička sustava, koja se koriste za pogon sljedećih komponenata: kočnica kotača (eng. *wheel brakes*), jedinice za upravljanje nosnom nogom po smjeru (eng. *nose wheel steering*), jedinice za uvlačenje/izvlačenje podvozja (eng. *landing gear retraction/extension*), zakrilca i predkrilca zrakoplova, sustava za kočenje motorom (eng. *thrust reversal*), spojlera i zračnih kočnica (eng.

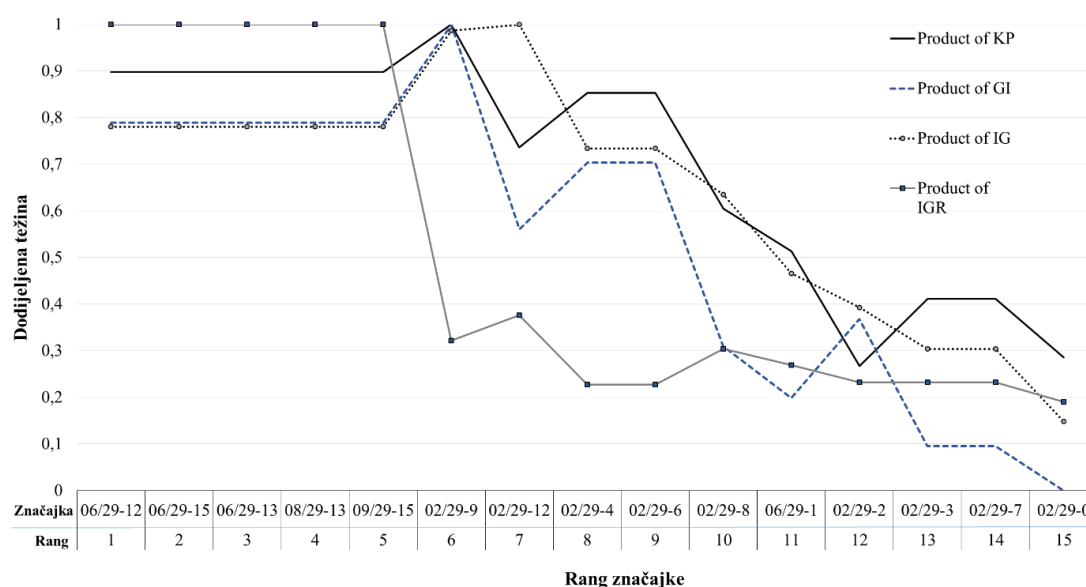
*spoilers/speed brakes*), kontrolnih površina (eng. *flight controls surfaces*), utovarnim vratima (eng. *cargo doors*), itd.

Hidraulički sustavi podložni su sljedećim kvarovima [104]:

- pregrijavanju sustava (eng. *system overheat*) – sustav prelazi maksimalnu dopuštenu radnu temperaturu i treba biti isključen,
- gubitak tlaka sustava (eng. *loss of system pressure*) – zbog gubitka hidrauličkog fluida ili pumpe,
- kontaminacija hidrauličkog fluida (eng. *hydraulic fluid contamination*) – kontaminacija može biti uzrokovana zbog nepravilnog održavanja hidrauličkog sustava ili kvara komponenata.

Posljedice ovih kvarova, tj. pregrijavanja sustava, gubitka tlaka ili onečišćenja fluida mogu rezultirati gubitkom hidrauličkog sustava i gubitkom funkcija onih komponenata koje ovaj sustav pogoni.

Kombinacija svih poruka upozorenja po fazama leta koje je generirao sustav ATA 29 u promatranom razdoblju nalazi se u prilogu D, a na slici 5-3 je prikazan rang za prvih 50% značajki. Značajke su rangirane prema srednjoj vrijednosti svih težina, dobivene nakon izračuna pojedinačnih težina temeljem KP, GI, IG i IGR filterskog postupka (v. odjeljak 4.1.2), za  $D_{KMIN}$  skup podataka ATA 29 sustava.



Slika 5-3. ATA 29 – prikaz ranga prvih 50% značajki

Iz slike 5-3 vidljiv je pravilan trend dodjeljivanja težina značajkama primjenom KP, GI i IG filterskih postupaka, dok su manja odstupanja vidljiva kod dodjeljivanja težina značajkama primjenom IGR filterskog postupka.

#### **5.2.1.4. ATA 32**

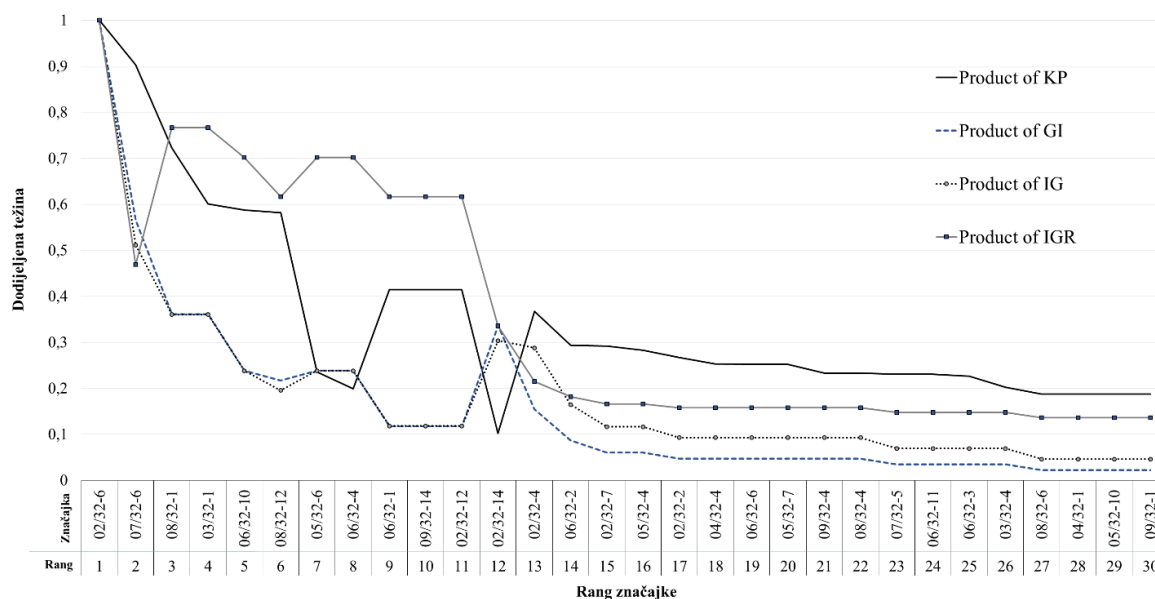
Sustav podvozja (ATA 32, eng. *landing gear*) je dio zrakoplova namijenjen za njegovo sigurno polijetanje, slijetanje i kretanje po tlu. Podvozje sačinjavaju glavne noge (eng. *main gear*) i nosna noga (eng. *nose gear*) koje se, nakon uzlijetanja, uvlače u trup.

Glavne noge su opremljene s ugljičnim kočnicama (eng. *carbon brakes*), sustavom protiv blokiranja kotača (eng. *anti-skid system*), automatskim sustavom za kočenje (eng. *automatic braking system*), jedinicom za hlađenje kočnica (eng. *brake cooling fan*) i sustavom za nadzor tlaka u gumama (eng. *tyre pressure indicating system*) [104].

Glavne noge i vrata glavnih nogu su kontrolirana pomoću dva LGCIU-a (eng. *Landing Gear Control and Interface Units*). Drugim riječima, glavne noge su kontrolirane elektronički (glavne funkcije kao izvlačenje i uvlačenje nogu), a pogonjene hidraulički (otvaranje vrata). LGCIU također prikuplja i obrađuje podatke vezane uz sigurnosne brave podvozja (eng. *gear locks*), amortizere (eng. *shock absorber struts*), vrata podvozja, utovarna vrata (eng. *cargo doors*), itd.. Podvozje, vrata podvozja, amortizeri i utovarna vrata opremljeni su senzorima (eng. *proximity detectors*) za nadzor i prikaz položaja tih komponenata [104].

Prednja (nosna) noga je opremljena sustavom za upravljanje nosnom nogom po smjeru (eng. *nose wheel steering system*). Sustav za upravljanje nosnom nogom po smjeru je upravljan elektronički pomoću BSCU računala (eng. *Brake and Steering Control Unit*), a pogonjen hidraulički. BSCU dobiva ulazne naredbe za upravljanje nosnom nogom po smjeru putem upravljačkog volana (eng. *steering handwheels*), autopilota i pedala za upravljanje kormilom pravca (eng. *rudder pedals*) [104].

Kombinacija svih poruka upozorenja po fazama leta koje je generirao sustav ATA 32 u promatranom razdoblju nalazi se u prilogu D, a na slici 5-4 je prikazan rang za prvih 50% značajki. Značajke su rangirane prema srednjoj vrijednosti svih težina, dobivene nakon izračuna pojedinačnih težina temeljem KP, GI, IG i IGR filterskog postupka (v. odjeljak 4.1.2), za  $D_{KMIN}$  skup podataka ATA 32 sustava.



Slika 5-4. ATA 32 – prikaz ranga prvih 50% značajki

Iz slike 5-4 vidljiv je pravilan trend dodjeljivanja težina značajkama kojima je rang niži od 13. Manja odstupanja kod dodjeljivanja težina značajkama su vidljiva kod početno rangiranih značajki.

#### 5.2.1.5. ATA 34

Navigacijski sustav (ATA 34, eng. *navigation*) se može podijeliti u više grupa sustava, kao što je ADIRS sustav (eng. *Air Data and Internal Reference System*), sustava radio navigacije (eng. *radio navigation*) te dodatnih navigacijskih sustava. Unutar svake grupe sustava nalazi se više podsustava.

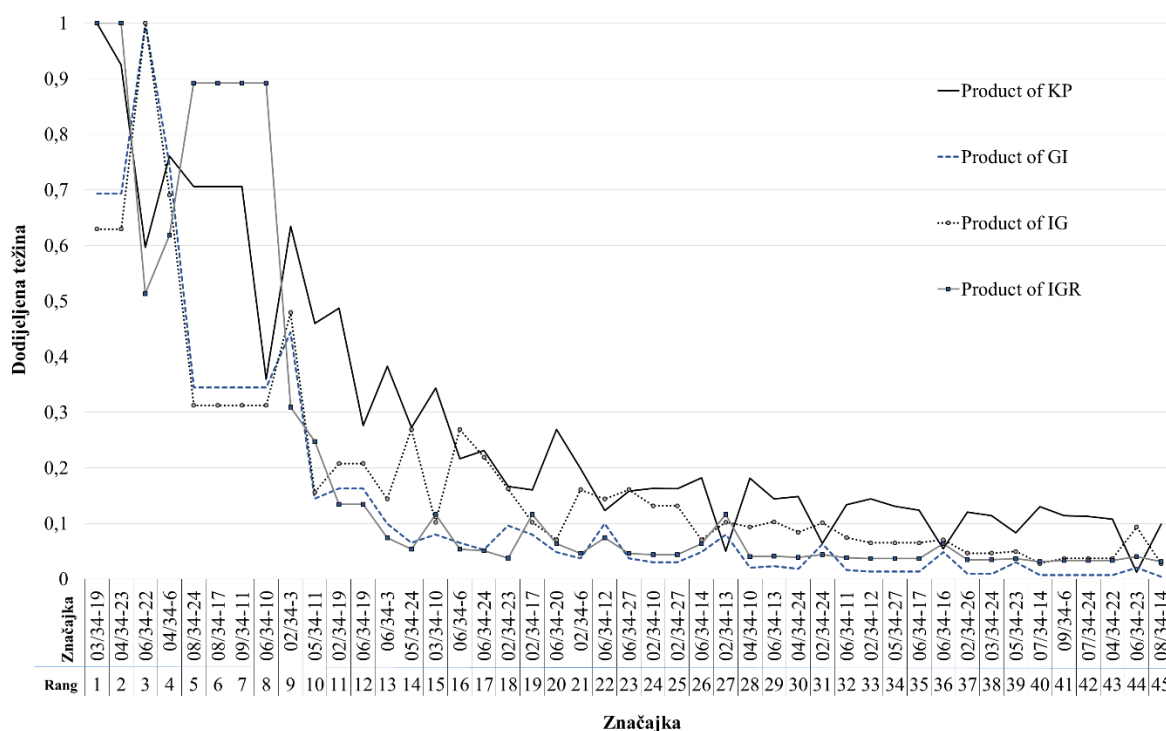
ADIRS sustav se sastoji od tri identične ADIRU jedinice (eng. *Air Data Internal Reference Units*). Svaka ADIRU jedinica je kombinacija ADR računala (eng. *Air Data Reference computer*) i IR sustava (eng. *Internal Reference*) baziranog na laserskom žiroskopu. ADR računalo sakuplja ulazne podatke iz sonda i senzora te daje informacije o aerodinamičkoj brzini (eng. *airspeed*), Mach broju (eng. *Mach number*), vanjskoj temperaturi, visini po tlaku (eng. *barometric altitude*), napadnom kutu zrakoplova i upozorenjima o prekoračenju brzine (eng. *overspeed warnings*). IR sustav daje informacije o smjeru (eng. *heading*), položaju zrakoplova u prostoru (stav, eng. *aircraft attitude*), smjeru brzine leta (eng. *track*), ubrzanju (eng. *acceleration*), brzini zrakoplova u odnosu na zemlju (eng. *ground speed*) [104].

Sustav radio navigacije sastoji se niz sustava kao što su GPS (eng. *Global Positioning System*), VOR sustava (eng. *VHF omnidirectional range*), ADF (eng. *automatic direction finder*), DME (eng. *distance measuring equipment*), itd.. U slučaju nepovoljnih vremenskih uvjeta (magla,

olujno nevrijeme, itd.), ILS sustav (eng. *Instrument Landing System*) omogućava sigurno slijetanje zrakoplova po optimalnoj putanji [104].

Pod dodatne navigacijske sustave spadaju sustavi kao što su sustav za upozoravanje na opasno približavanje zemlji (eng. *Enhanced Ground Proximity Warning System*, EGPWS), radar za prikazivanje trenutne vremenske slike (eng. *weather radar*), itd [104].

Kombinacija svih poruka upozorenja po fazama leta koje je generirao sustav ATA 34 u promatranom razdoblju nalazi se u prilogu D, a na slici 5-5 je prikazan rang za prvih 50% značajki. Značajke su rangirane prema srednjoj vrijednosti svih težina, dobivene nakon izračuna pojedinačnih težina temeljem KP, GI, IG i IGR filterskog postupka (v. odjeljak 4.1.2), za  $D_{KMIN}$  skup podataka ATA 34 sustava.



Slika 5-5. ATA 34 – prikaz ranga prvih 50% značajki

Iz slike 5-5 vidljiv je pravilan trend dodjeljivanja težina značajkama kojima je rang niži od 10, dok su manja odstupanja vidljiva kod dodjeljivanja težina značajkama s većim rangom.

### 5.2.2. Rezultati asocijativnih pravila

Proces za pronalaženje asocijativnih pravila, razvijen pomoću RapidMiner platforme, rezultirao je otkrivanjem asocijativnih pravila jedino za sustav ATA 29. Unutar  $D_{KMIN}$  skupova podataka preostalih sustava (ATA 22, ATA 27, ATA 32 i ATA 34) nisu pronađena asocijativna pravila

koja bi zadovoljila početni kriterij minimalne podrške, tj. da se dvije značajke pojavljuju u minimalno 10% uzoraka unutar skupa podataka.

U tablici 5-1 prikazana su otkrivena asocijativna pravila unutar skupa podataka  $D_{KMIN}$  sustava ATA 29. Navedena pravila su zadovoljila kriterij minimalne podrške od 10%, a zatim i kriterij minimalne pouzdanosti od 50%. Osim asocijativnih pravila, u tablici 5-1 su prikazane i vrijednosti koeficijenta podrške i pouzdanosti.

Tablica 5-1. *Asocijativna pravila za značajku „06/29-12“*

A	06/29-12 (HYD Y ENG 2 PUMP LO PR)		
	Značajka	Podrška $P(A \cup B)$ [%]	Pouzdanost $P(B   A)$ [%]
B	06/29-13 (HYD Y RSVR LO LVLXX)	37,50	75,00
	06/29-15 (HYD Y SYS LO PR)	37,50	75,00
	09/29-15 (HYD Y SYS LO PR)	37,50	75,00

Otkrivena pravila prikazana u tablici 5-1 označavaju slučaj kvara kada je došlo do propuštanja ulja u hidrauličkom sustavu zbog puknuća cijevi. U tom slučaju se indicira upozorenje da je u rezervoaru niska razina fluida (HYD Y RSVR LO LVLXX) te nakon što fluid iscure, hidraulička pumpa ostane bez ulja, zrak uđe u sustav i to rezultira upozorenjem da je niski pritisak u sustavu (HYD Y ENG 2 PUMP LO PR, HYD Y SYS LO PR).

### 5.3. Rezultati klasifikacijskih modela prve grupe

Kao što je prikazano u odjeljku 4.2, ulaz u proces za izgradnju klasifikacijskih modela prve grupe je skup podataka  $D_{KMI-1}$  koji obuhvaća sjedinjene zapise o porukama upozorenja (WN) i prošlim pilotskim defektima (PI-DE) te na temelju ulaznih značajki (kombinacije šifriranih poruka upozorenja po fazama leta) predviđa zavisnu značajku *Predviđa li se zapis na sljedećem letu?*

#### 5.3.1. ATA 22

Transformacijom skupa podataka  $D_{KMI}$  u skup  $D_{KMI-1}$  (v. odjeljak 4.2.1), grupirani su svi uzorci koji su zabilježeni na različitim letovima na isti datum, a da nisu rezultirali upisom u TLB. Analogno tome, broj uzoraka u skupu  $D_{KMI-1}$  sustava ATA 22 za zrakoplove  $Ac_1$ ,  $Ac_2$ ,  $Ac_3$  i  $Ac_4$  je smanjen u odnosu na broj uzoraka u skupu  $D_{KMI}$  (slika 3-17). Zrakoplov  $Ac_1$  u skupu  $D_{KMI-1}$  bilježi 365 uzoraka ( $D_{KMI} = 396$  uzoraka),  $Ac_2$  417 uzoraka ( $D_{KMI} = 468$  uzoraka),  $Ac_3$  515 uzoraka ( $D_{KMI} = 622$  uzoraka), a  $Ac_4$  680 uzoraka ( $D_{KMI} = 835$  uzoraka).



### 5.3.1.1. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_1$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_1$  zabilježeno je 365 uzoraka, od čega je 348 uzoraka pripalo klasi 0, dok je preostalih 17 pripalo klasi 1. Navedeni skup je dalje podijeljen u skup za učenje modela (70%) i skup za verifikaciju (30%). Primjerci su dalje odabirani različitim postupcima uzorkovanja podataka (stratificiranom ili slučajnom) i dobivene su vrijednosti evaluacijskih mjera na temelju jedinstvenih primjeraka iz oba skupa.

Rezultati klasifikacijskog modela izgrađenog bez primjene filterskih postupaka i s ukupnim brojem ulaznih značajki za sustav ATA 22 (54 značajke, v. prilog D) i zrakoplov  $Ac_1$  su prikazani u tablici 5-2 za sve jedinstvene primjerke.

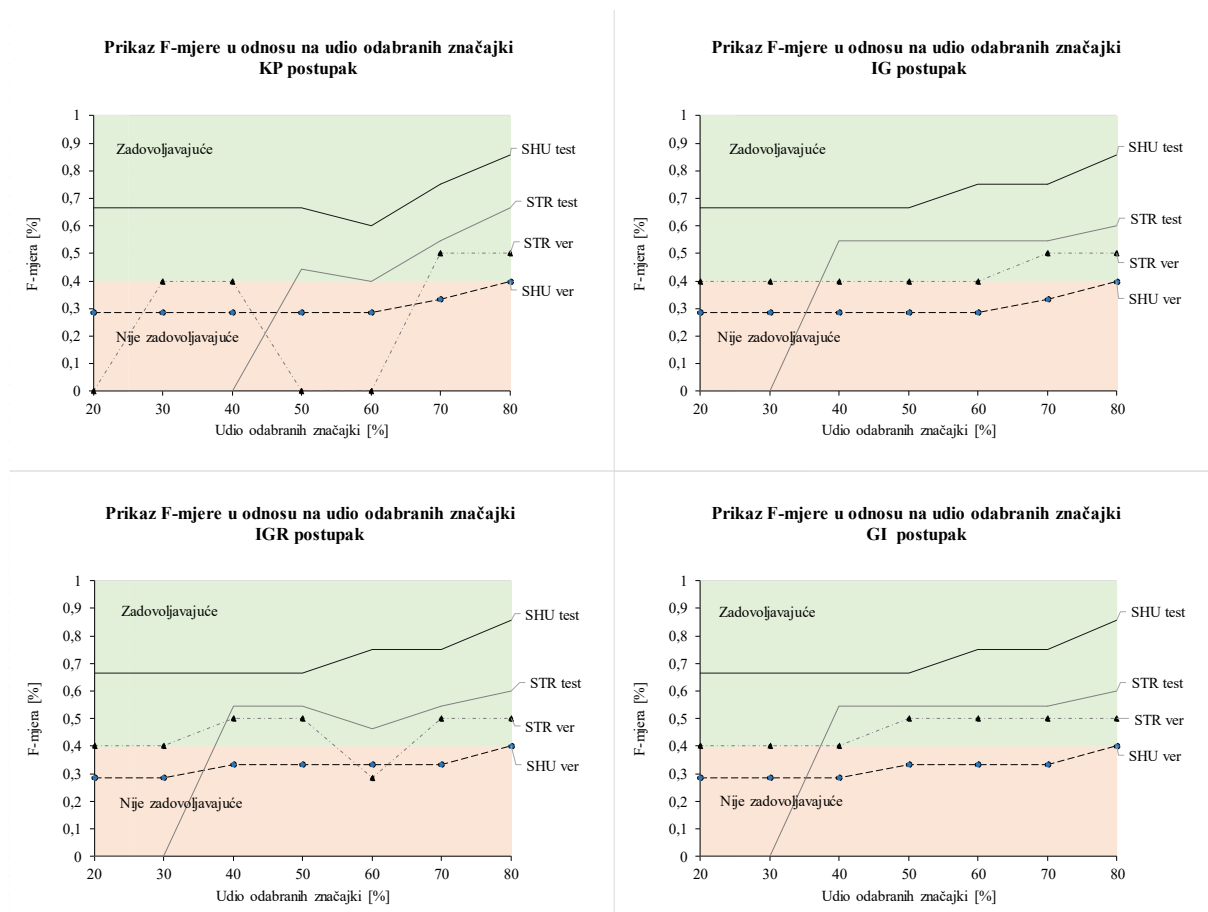
Tablica 5-2. ATA 22 zrakoplov  $Ac_1$  - vrijednosti  $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka

	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,6667	<b>0,4000</b>	0,6667	<b>0,4000</b>
<b>Osjetljivost*</b>	1,0000	<b>1,0000</b>	1,0000	<b>1,0000</b>
<b>Specifičnost**</b>	0,8333	<b>0,8000</b>	0,8500	<b>0,7500</b>

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-2 je vidljivo kako klasifikacijski model izgrađen stratificiranim uzorkovanjem podataka (STR) rezultira gotovo jednakim vrijednostima statističkih mjera kao i model izgrađen slučajnim uzorkovanjem podataka (SHU). Vrijednosti  $F$ -mjere i osjetljivosti jednake za oba modela, no vrijednost mjere specifičnost je neznatno bolja u slučaju modela izgrađenim slučajnim uzorkovanjem podataka. Temeljem prikazanih vrijednosti statističkih mjera, smatra se kako su oba modela reprezentativna. Slijedeći uvodne napomene u ovom poglavlju, ispitat će se mogućnosti poboljšanja ovih mjera smanjenjem broja ulaznih značajki. U slučaju  $F$ -mjere prag je postavljen na 0,4 (40%), tj. ukoliko će model s manjim brojem značajki rezultirati boljim vrijednostima mjera, smatra se da taj model ima bolje generalizacijske sposobnosti za klasifikaciju novih primjeraka. Na slici 5-6 prikazane su vrijednosti  $F$ -mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano rang značajki  $R_{KMIi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ . Vrijednosti mjera osjetljivosti i specifičnosti za svaki od novoizgrađenih modela su prikazane u tablicama 5-3 i 5-4.



Slika 5-6. ATA 22 – prikaz vrijednosti  $F$ -mjere za  $Ac_1$

Iz slike 5-6 je vidljivo kako primjenom ranga značajki dobivenog KP postupkom na verifikacijskom skupu podataka (oznaka *ver*), kada je udio ulaznih značajki smanjen na 70% i 80%, modeli izgrađeni stratificiranim uzorkovanjem podataka rezultiraju boljim vrijednostima  $F$ -mjere (0,5000). Isti zaključak vrijedi i za rang značajki dobiven IG filterskim postupkom.

Također je vidljivo da su bolje vrijednosti  $F$ -mjere postignute kada je udio ulaznih značajki smanjen na 40%, 50%, 70% i 80% temeljem IGR postupka i model izgrađen stratificiranim uzorkovanjem podataka.

Bolje vrijednosti  $F$ -mjere su postignute i kada je udio ulaznih značajki smanjen na 50%, 60%, 70% i 80% temeljem GI filterskog postupka, a modeli izgrađeni stratificiranim uzorkovanjem podataka.

Iako su iste vrijednost  $F$ -mjere postignute već za slučaj kada je model izgrađen stratificiranim uzorkovanjem podataka, a udio ulaznih značajki je smanjen na 20% neovisno o primijenjenom filterskom postupku, vrijednosti  $F$ -mjere na testnom skupu podataka su loše te ti modeli nisu reprezentativni.

Tablica 5-3. ATA 22 – prikaz vrijednosti mjere osjetljivosti za  $Ac_1$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,7500	<b>1,0000</b>	0,0000	0,0000	0,7500	<b>1,0000</b>	0,0000	1,0000	0,7500	<b>1,0000</b>	0,0000	1,0000	0,7500	<b>1,0000</b>	0,0000	1,0000
30	0,7500	<b>1,0000</b>	0,0000	1,0000	0,7500	<b>1,0000</b>	0,0000	1,0000	0,7500	<b>1,0000</b>	0,0000	1,0000	0,7500	<b>1,0000</b>	0,0000	1,0000
40	0,7500	<b>1,0000</b>	0,0000	1,0000	0,7500	<b>1,0000</b>	0,7500	<b>1,0000</b>	0,7500	<b>1,0000</b>	0,7500	<b>1,0000</b>	0,7500	<b>1,0000</b>	0,7500	<b>1,0000</b>
50	0,7500	<b>1,0000</b>	0,5000	0,0000	0,7500	<b>1,0000</b>	0,7500	<b>1,0000</b>	0,7500	<b>1,0000</b>	0,7500	<b>1,0000</b>	0,7500	<b>1,0000</b>	0,7500	<b>1,0000</b>
60	0,7500	<b>1,0000</b>	0,5000	0,0000	1,0000	<b>1,0000</b>	0,7500	<b>1,0000</b>	1,0000	<b>1,0000</b>	0,7500	<b>1,0000</b>	1,0000	<b>1,0000</b>	0,7500	<b>1,0000</b>
70	1,0000	<b>1,0000</b>	0,7500	<b>1,0000</b>	1,0000	<b>1,0000</b>	0,7500	<b>1,0000</b>	1,0000	<b>1,0000</b>	0,7500	<b>1,0000</b>	1,0000	<b>1,0000</b>	0,7500	<b>1,0000</b>
80	1,0000	<b>1,0000</b>	0,7500	<b>1,0000</b>	1,0000	<b>1,0000</b>	0,7500	<b>1,0000</b>	1,0000	<b>1,0000</b>	0,7500	<b>1,0000</b>	1,0000	<b>1,0000</b>	0,7500	<b>1,0000</b>

Tablica 5-4. ATA 22 – prikaz vrijednosti mjere specifičnosti za  $Ac_1$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,9417	0,6667	0,9083	1,0000	0,9417	0,6667	0,9000	0,7857	0,9417	0,6667	0,9000	0,7500	0,9417	0,6667	0,9000	0,7500
30	0,9250	0,6667	0,8519	0,7500	0,9417	0,6667	0,8583	0,7500	0,9417	0,6667	0,8621	0,7500	0,9417	0,6667	0,8583	0,7500
40	0,9250	0,6667	0,8462	0,7500	0,9417	0,7333	0,8333	<b>0,8333</b>	0,9250	0,6667	0,8333	0,7500	0,9250	0,6667	0,8417	0,7500
50	0,9000	0,6429	0,9000	<b>0,9167</b>	0,9250	0,7333	0,8333	<b>0,8333</b>	0,9250	0,6667	0,8417	0,7500	0,9250	0,7333	0,8333	<b>0,8333</b>
60	0,8000	0,6429	0,8500	<b>0,9167</b>	0,8500	0,7143	0,6833	0,5833	0,8667	0,6429	0,8333	0,7500	0,8500	0,7333	0,8167	<b>0,8333</b>
70	0,8500	0,7143	0,8167	<b>0,8333</b>	0,8500	0,7143	0,8167	<b>0,8333</b>	0,8500	0,7143	0,8667	<b>0,8333</b>	0,8500	0,7143	0,8667	<b>0,8333</b>
80	0,9500	0,7857	0,9000	<b>0,8333</b>	0,9500	0,7857	0,8500	<b>0,8333</b>	0,9500	0,7857	0,8500	<b>0,8333</b>	0,9500	0,7857	0,8500	<b>0,8333</b>

Unutar tablica 5-3 i 5-4 podebljani su rezultati mjera *osjetljivosti* i *specifičnosti* koji prikazuju iste ili bolje vrijednosti u odnosu na vrijednosti dobivene modelom koji je izgrađen s ukupnim brojem ulaznih značajki (tablica 5-2).

Promatranjem vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-3 je vidljivo kako gotovo svi izgrađeni modeli slučajnim uzorkovanjem podataka (oznaka *SHU ver*), neovisno o udjelu ulaznih značajki i filterskom postupku, rezultiraju maksimalnom vrijednošću mjere *osjetljivost*, što znači da će svi primjerci iz pozitivne klase biti točno označeni.

Vrijednosti mjere *specifičnosti* unutar tablice 5-4 ukazuju da će uvjetna vjerojatnost ispravno klasificiranih negativnih primjeraka biti uvijek veća u slučaju kada je model izgrađen stratificiranim uzorkovanjem podataka (oznaka *STR ver*) i dok je udio ulaznih značajki smanjen na 70% i 80% neovisno o filterskom postupku.

### 5.3.1.2. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_2$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_2$  zabilježeno je 417 uzoraka, od čega je 397 uzoraka pripalo klasi 0, dok je preostalih 20 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 292 uzoraka, od kojih je 14 svrstano u klasu 1, a 278 u klasu 0. U skupu za verifikaciju (30%) svrstano je 125 uzoraka, od kojih je 119 pripalo klasi 0, a 6 klasi 1.

Rezultati klasifikacijskog modela izgrađenog bez primjene s ukupnim brojem ulaznih značajki za sustav ATA 22 su prikazani u tablici 5-5.

Tablica 5-5. ATA 22 zrakoplov  $Ac_2$  - vrijednosti  $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka

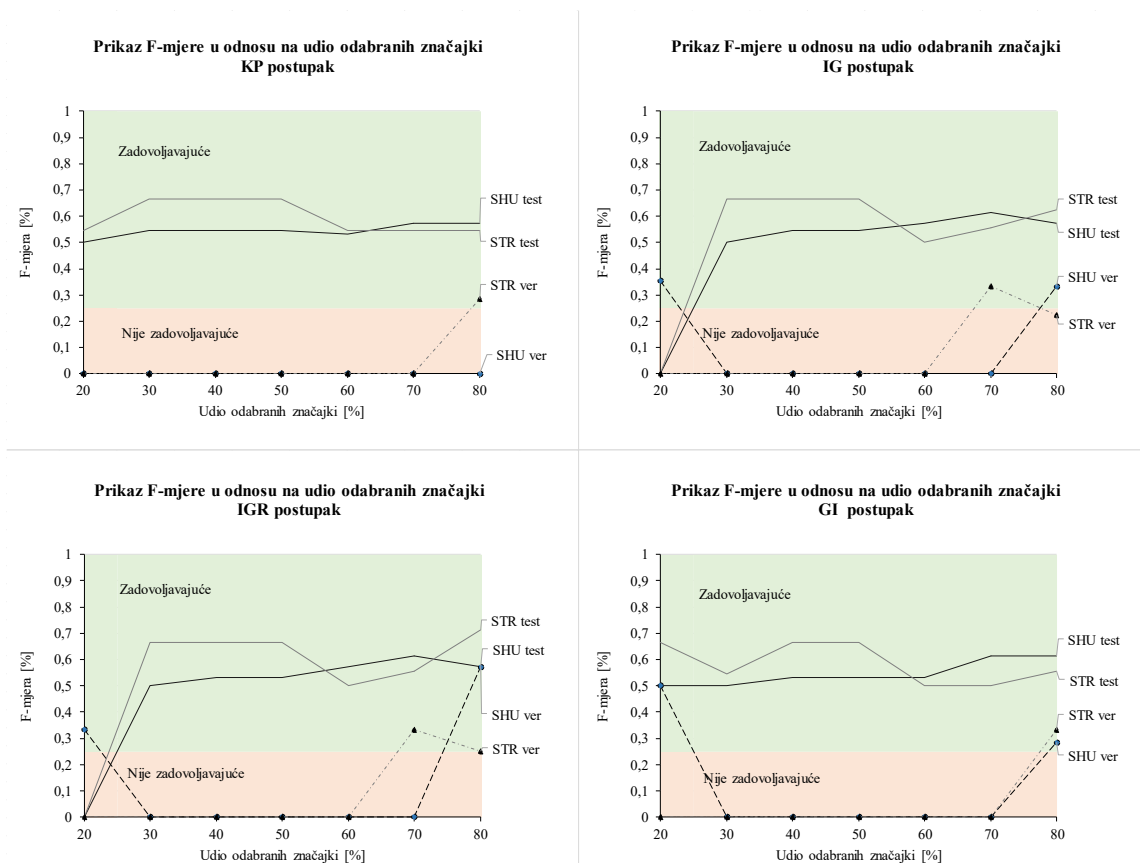
	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,5000	0,0000	0,7143	<b>0,2500</b>
<b>Osjetljivost*</b>	0,3333	0,0000	0,8333	<b>0,3333</b>
<b>Specifičnost**</b>	1,0000	1,0000	0,9300	<b>0,7647</b>

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-5 je vidljivo kako klasifikacijski model izgrađen stratificiranim uzorkovanjem podataka (STR) rezultira boljim vrijednostima statističkih mjera od modela izgrađenog slučajnim uzorkovanjem podataka (SHU). Niske vrijednosti  $F$ -mjere i *osjetljivosti* upućuju na to da u ovom skupu podataka postoji veća sličnost među uzorcima koji pripadaju klasi 1 i klasi 0. Na

slici 5-7 prikazane su vrijednosti  $F$ -mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano g ranga značajki  $R_{KMIi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ . Prag za  $F$ -mjeru je postavljen na 0,25 (25%). Vrijednosti mjera osjetljivosti i specifičnosti za svaki od novoizgrađenih modela su prikazane u tablicama 5-6 i 5-7.



Slika 5-7. ATA 22 – prikaz vrijednosti  $F$ -mjere za  $Ac_2$

Iz slike 5-7 je vidljivo kako primjenom ranga značajki dobivenog IGR filterskim postupkom, kada je udio ulaznih značajki smanjen na 80%, model izgrađen slučajnim uzorkovanjem podataka rezultira boljim vrijednostima  $F$ -mjere i smatra se da taj model ima bolje generalizacijske sposobnosti za klasifikaciju novih primjeraka od modela izgrađenog s ukupnim brojem ulaznih značajki. Isti zaključci se mogu izvesti i za slučaj kada je udio značajki smanjen na 20% temeljem GI filterskim postupkom, a model izgrađen slučajnim uzorkovanjem.

Boljim vrijednostima  $F$ -mjere rezultiraju i modeli izgrađeni stratificiranim uzorkovanjem podataka kada je udio značajki smanjen na 70% temeljem IGR i IG filterskog postupka.

Tablica 5-6. ATA 22 – prikaz vrijednosti mjere osjetljivosti za  $Ac_2$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,3333	0,0000	0,5000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,3333	0,3333	0,5000	0,0000
30	0,5000	0,0000	0,5000	0,0000	0,3333	0,0000	0,5000	0,0000	0,3333	0,0000	0,5000	0,0000	0,3333	0,0000	0,5000	0,0000
40	0,5000	0,0000	0,5000	0,0000	0,6667	0,0000	0,5000	0,0000	0,5000	0,0000	0,5000	0,0000	0,6667	0,0000	0,5000	0,0000
50	0,5000	0,0000	0,5000	0,0000	0,6667	0,0000	0,5000	0,0000	0,5000	0,0000	0,5000	0,0000	0,6667	0,0000	0,5000	0,0000
60	0,6667	0,0000	0,5000	0,0000	0,6667	0,0000	0,3333	0,0000	0,6667	0,0000	0,3333	0,0000	0,6667	0,0000	0,3333	0,0000
70	0,6667	0,0000	0,5000	0,0000	0,6667	0,0000	0,8333	<b>0,6667</b>	0,6667	0,0000	0,8333	<b>0,6667</b>	0,6667	0,0000	0,3333	0,0000
80	0,6667	0,0000	0,5000	<b>0,6667</b>	0,6667	<b>0,6667</b>	0,8333	0,3333	0,6667	0,3333	0,8333	0,3333	0,6667	0,3333	0,8333	<b>0,6667</b>

Tablica 5-7. ATA 22 – prikaz vrijednosti mjere specifičnosti za  $Ac_2$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	1,0000	<b>1,0000</b>	0,9500	<b>1,0000</b>	0,6683	0,2941	0,8850	<b>1,0000</b>	0,6933	0,3529	0,7850	<b>1,0000</b>	1,0000	<b>1,0000</b>	1,0000	<b>1,0000</b>
30	0,9500	<b>0,9412</b>	1,0000	<b>1,0000</b>	1,0000	<b>1,0000</b>	1,0000	<b>1,0000</b>	1,0000	<b>1,0000</b>	1,0000	<b>1,0000</b>	1,0000	<b>1,0000</b>	0,9500	<b>1,0000</b>
40	0,9500	<b>0,9412</b>	1,0000	<b>1,0000</b>	0,8800	<b>0,9412</b>	1,0000	<b>1,0000</b>	0,9500	<b>0,9412</b>	1,0000	<b>1,0000</b>	0,8800	<b>0,9412</b>	1,0000	<b>1,0000</b>
50	0,9500	<b>0,9412</b>	1,0000	<b>1,0000</b>	0,8800	<b>0,9412</b>	1,0000	<b>1,0000</b>	0,9500	<b>0,9412</b>	1,0000	<b>1,0000</b>	0,8800	<b>0,9412</b>	1,0000	<b>1,0000</b>
60	0,8717	<b>0,9412</b>	0,9417	<b>1,0000</b>	0,9050	<b>0,9412</b>	1,0000	<b>1,0000</b>	0,9050	<b>0,9412</b>	1,0000	<b>1,0000</b>	0,8717	<b>0,9412</b>	1,0000	<b>1,0000</b>
70	0,9050	<b>0,9412</b>	0,9500	<b>1,0000</b>	0,9300	<b>0,9412</b>	0,8317	0,5882	0,9300	<b>0,9412</b>	0,8083	0,5882	0,9300	<b>0,9412</b>	1,0000	<b>1,0000</b>
80	0,9050	<b>0,9412</b>	0,9500	0,4706	0,8967	<b>0,8824</b>	0,9167	0,7647	0,8967	<b>0,8824</b>	0,8717	0,7059	0,9300	<b>0,8235</b>	0,8317	0,5882

Promatranjem vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-6 je vidljivo da će vjerojatnost ispravno klasificiranih pozitivnih primjeraka biti dobra za slučaj kada je model izgrađen stratificiranim uzorkovanjem podataka, a udio ulaznih značajki smanjen na 80% temeljem KP i IG filterskih postupaka i 70% temeljem IGR i IG filterskih postupaka. Iste vrijednosti te mjere su dobivene kada je model izgrađen slučajnim uzorkovanjem podataka, a udio ulaznih značajki smanjen na 80% temeljem IGR postupka.

Vrijednosti mjere *specifičnosti* unutar tablice 5-7 ukazuju da će uvjetna vjerojatnost ispravno klasificiranih negativnih primjeraka biti uvijek veća u slučaju kada je model izgrađen slučajnim uzorkovanjem podataka neovisno o filterskom postupku i udjelu ulaznih značajki.

### 5.3.1.3. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_3$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_3$  zabilježeno je 515 uzoraka, od čega je 491 uzorak pripao klasi 0, a preostala 24 uzoraka su pripala klasi 1. Skupu za učenje modela (70%) pripalo je 361 uzoraka, od kojih je 17 svrstano u klasu 1, a 344 u klasu 0. U skupu za verifikaciju (30%) svrstano je 154 uzoraka, od kojih je 147 pripalo klasi 0, a 7 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki za sustav ATA 22 su prikazani u tablici 5-8.

Tablica 5-8. ATA 22 zrakoplov  $Ac_3$  - vrijednosti *F*-mjere, *osjetljivosti* i *specifičnosti* za SHU i STR uzorkovanje podataka

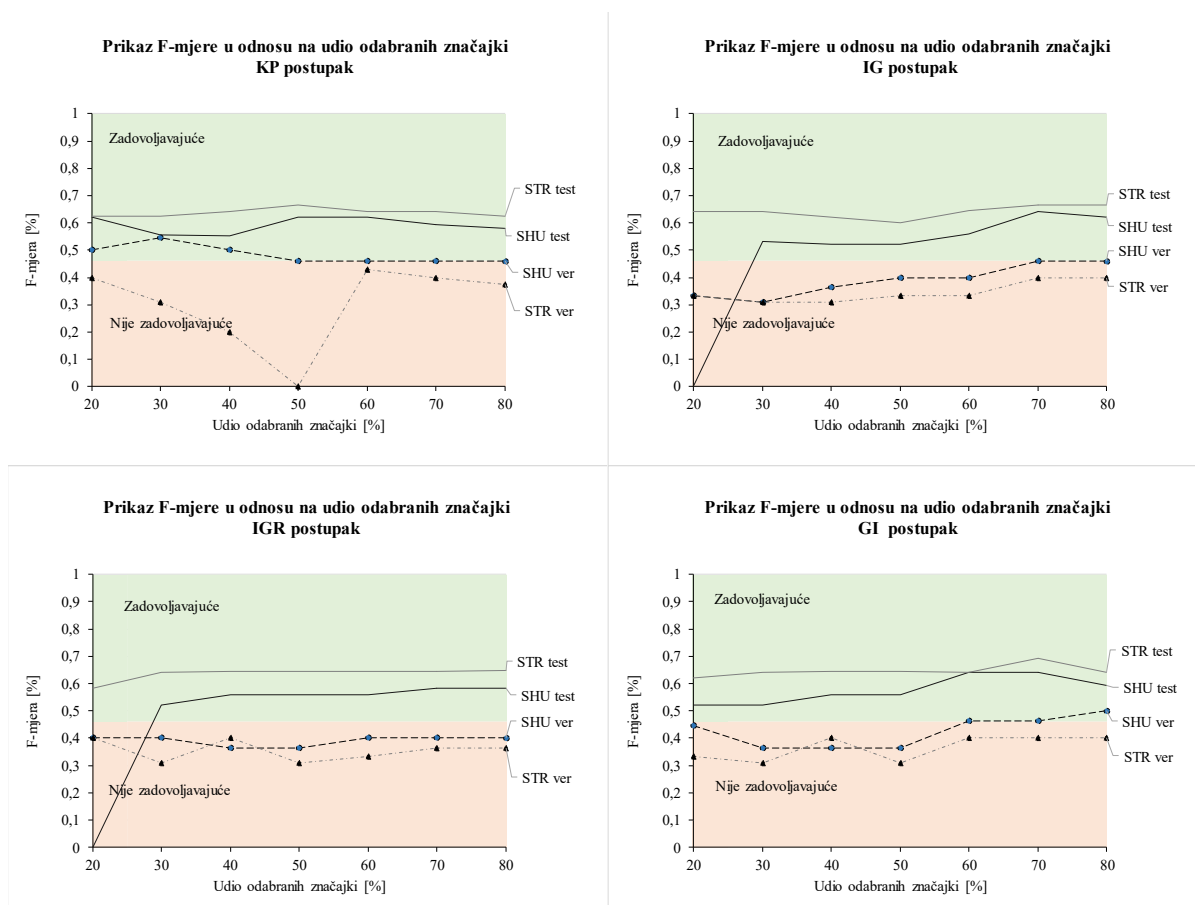
	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,6000	<b>0,4615</b>	0,6667	0,4000
<b>Osjetljivost*</b>	0,6923	<b>0,6000</b>	0,6923	0,3333
<b>Specifičnost**</b>	0,8738	<b>0,8214</b>	0,9157	0,9259

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-8 je vidljivo kako klasifikacijski model izgrađen slučajnim uzorkovanjem podataka (SHU) ima bolje vrijednosti *F*-mjere i *osjetljivosti* u odnosu na model izgrađen stratificiranim uzorkovanjem podataka (STR). Iako je mjera *specifičnosti* bolja za slučaj kada je model izgrađen nad skupom podataka koji je stratificirano uzorkovan, zbog boljih vrijednosti *F*-mjere i *osjetljivosti*, model izgrađen slučajnim uzorkovanjem je uzet kao reprezentativan. Na slici 5-8 prikazane su vrijednosti *F*-mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne

značajke u udjelu od 20% do 80% temeljem učitano­g ranga značajki  $R_{KMIi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ . Prag za  $F$ -mjeru je postavljen na 0,46 (46%). Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki od novoizgrađenih modela su prikazane u tablicama 5-9 i 5-10.



Slika 5-8. ATA 22 – prikaz vrijednosti  $F$ -mjere za  $Ac_3$

Iz slike 5-8 je vidljivo kako vrijednosti  $F$ -mjere iznad zadanog praga daju modeli izgrađeni slučajnim uzorkovanjem podataka kada je udio ulaznih značajki smanjen na 20% i 30% temeljem KP filterskog postupka i kada je udio značajki smanjen na 80% temeljem GI postupka.

Iste vrijednost  $F$ -mjere su postignute za slučaj modela izgrađenih slučajnim uzorkovanjem kada je udio ulaznih značajki smanjen od 50% do 80% temeljem KP postupka, 70% i 80% temeljem IG postupka te 60% i 70% temeljem GI postupka.



Tablica 5-9. ATA 22 – prikaz vrijednosti mjere osjetljivosti za  $Ac_3$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,6923	<b>0,6000</b>	0,7692	0,3333	0,0000	0,4000	0,5385	0,3333	0,0000	0,2000	0,6923	0,3333	0,4615	0,4000	0,6923	0,3333
30	0,7143	<b>0,6000</b>	0,7692	0,3333	0,4615	0,4000	0,6923	0,3333	0,5714	0,4000	0,6923	0,3333	0,4615	0,4000	0,6923	0,3333
40	0,6154	<b>0,6000</b>	0,6923	0,1667	0,5385	0,4000	0,7692	0,5000	0,4615	0,4000	0,6923	0,3333	0,5385	0,4000	0,7692	0,5000
50	0,6923	<b>0,6000</b>	0,7692	0,0000	0,5385	0,4000	0,7692	0,3333	0,4615	0,4000	0,6923	0,3333	0,5385	0,4000	0,7692	0,3333
60	0,6923	<b>0,6000</b>	0,6154	0,5000	0,5385	0,4000	0,7692	0,3333	0,5385	0,4000	0,7692	0,3333	0,6923	<b>0,6000</b>	0,6154	0,3333
70	0,6154	<b>0,6000</b>	0,6154	0,3333	0,5385	0,4000	0,7692	0,3333	0,6923	<b>0,6000</b>	0,6154	0,3333	0,6923	<b>0,6000</b>	0,6923	0,3333
80	0,6923	<b>0,6000</b>	0,7692	0,5000	0,5385	0,4000	0,8462	0,3333	0,6923	<b>0,6000</b>	0,6154	0,3333	0,6154	<b>0,6000</b>	0,6923	0,3333

Tablica 5-10. ATA 22 – prikaz vrijednosti mjere specifičnosti za  $Ac_3$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,8790	0,8571	0,8610	0,9259	0,9024	0,8929	0,9357	0,9259	0,9405	0,8571	0,8981	0,8519	0,9381	0,9286	0,8905	0,8519
30	0,8062	0,8929	0,8667	0,8148	0,9381	0,8929	0,9071	0,8148	0,8681	0,7857	0,9014	0,8148	0,9405	0,8571	0,9024	0,8148
40	0,8738	0,8571	0,9038	0,8889	0,9181	0,8571	0,8786	0,7778	0,9381	0,8571	0,8871	0,8148	0,9181	0,8571	0,8786	0,7778
50	0,8881	0,8214	0,8929	0,0000	0,9181	0,8571	0,8729	0,8148	0,9381	0,8929	0,8762	0,8519	0,9181	0,8571	0,8729	0,8148
60	0,8881	0,8214	0,9348	0,8148	0,9181	0,8929	0,8729	0,8519	0,9181	0,8929	0,8729	0,8519	0,9048	0,8214	0,9300	0,9259
70	0,9048	0,8214	0,9333	0,9259	0,9357	0,8929	0,8729	0,8889	0,9048	0,8214	0,9524	0,9259	0,9048	0,8214	0,9381	0,9259
80	0,8552	0,8214	0,8610	0,7407	0,9357	0,8929	0,8524	0,8889	0,8881	0,8214	0,9524	0,9259	0,9048	0,8571	0,9014	0,9259

Promatranjem vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-9 je vidljivo kako smanjenjem udjela ulaznih značajki bez obzira na filterski postupak i način uzorkovanja podataka, prag od 0,6 (60%) za mjeru osjetljivosti nije prijedn. Iste vrijednosti mjere *osjetljivosti* su postignute za slučaj modela izgrađenih slučajnim uzorkovanjem podataka bez obzira na udio odabranih značajki temeljem KP filterskog postupka, sa smanjenim udjelom značajki na 70% i 80% temeljem IG postupka te 60%, 70% i 80% temeljem GI postupka.

Vrijednosti mjere *specifičnosti* unutar tablice 5-10 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih modela izgrađenih neovisno o filterskom postupku i udjelu ulaznih značajki.

#### 5.3.1.4. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_4$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_4$  zabilježeno je 680 uzoraka, od čega je 649 uzoraka pripalo klasi 0, dok je preostalih 31 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 476 uzoraka, od kojih je 27 svrstano u klasu 1, a 449 u klasu 0. U skupu za verifikaciju (30%) svrstano je 204 uzoraka, od kojih je 200 pripalo klasi 0, a 4 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-11.

Tablica 5-11. *ATA 22 zrakoplov  $Ac_4$  - vrijednosti  $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka*

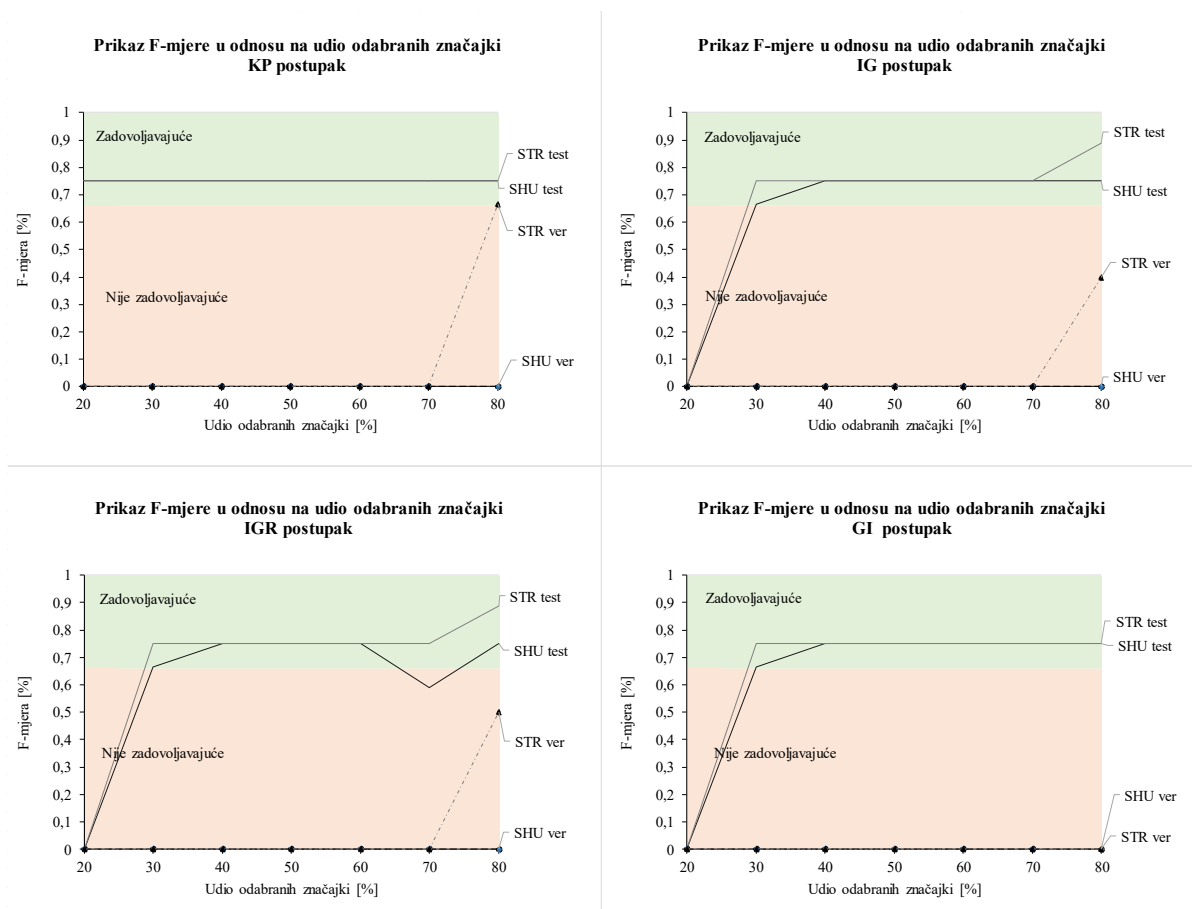
	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,8571	<b>0,6667</b>	0,8889	0,4000
<b>Osjetljivost*</b>	0,7500	<b>0,5000</b>	0,8000	0,5000
<b>Specifičnost**</b>	1,0000	<b>0,9600</b>	1,0000	0,9310

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-46 je vidljivo kako klasifikacijski model izgrađen slučajnim uzorkovanjem podataka (SHU) ima bolje vrijednosti svih statističkih mjera u odnosu na model izgrađen stratificiranim uzorkovanjem podataka (STR). Stoga je model izgrađen slučajnim uzorkovanjem uzet kao reprezentativan. Na slici 5-9 prikazane su vrijednosti  $F$ -mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano g ranga značajki  $R_{KMIi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ . Prag za  $F$ -mjeru je postavljen na 0,66 (66%).

Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki od novoizgrađenih modela su prikazane u tablicama 5-12 i 5-13.



Slika 5-9. ATA 22 – prikaz vrijednosti  $F$ -mjere za  $Ac_4$

Iz slike 5-9 vidljivo je kako smanjenjem udjela ulaznih značajki neovisno o filterskom postupku i načinu uzorkovanja podataka, vrijednosti  $F$ -mjere neće prijeći zadani prag na verifikacijskom skupu podataka.

Također je vidljivo kako istu vrijednost  $F$ -mjere postiže model izgrađen stratificiranjem uzorkovanjem podataka kada je udio ulaznih značajki smanjen na 80% temeljem KP filterskog postupka.

Tablica 5-12. ATA 22 – prikaz vrijednosti mjere osjetljivosti za  $Ac_4$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,6000	0,0000	0,6000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
30	0,6000	0,0000	0,6000	0,0000	0,5000	0,0000	0,6000	0,0000	0,5000	0,0000	0,6000	0,0000	0,5000	0,0000	0,6000	0,0000
40	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000
50	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000
60	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000
70	0,6000	0,0000	0,6000	0,0000	0,5000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000	0,6000	0,0000
80	0,6000	0,0000	0,6000	<b>0,5000</b>	0,6000	0,0000	0,8000	<b>0,5000</b>	0,6000	0,0000	0,8000	<b>0,5000</b>	0,6000	0,0000	0,6000	0,0000

Tablica 5-13. ATA 22 – prikaz vrijednosti mjere specifičnosti za  $Ac_4$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
30	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
40	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
50	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9643	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
60	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
70	1,0000	1,0000	1,0000	1,0000	0,9732	0,9615	1,0000	1,0000	1,0000	0,9615	1,0000	1,0000	1,0000	0,9615	1,0000	1,0000
80	1,0000	1,0000	1,0000	1,0000	1,0000	0,9615	1,0000	0,9615	1,0000	0,9615	1,0000	0,9259	1,0000	0,9600	1,0000	1,0000

Promatranjem vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-12 je vidljivo kako smanjenjem udjela ulaznih značajki bez obzira na primijenjeni filterski postupak i način uzorkovanja podataka, prag od 0,5 (50%) za mjeru osjetljivosti dobiven modelom izgrađenim nad ukupnim brojem značajki (tablica 5-11) nije prijeđen.

Vrijednosti mjere *specifičnosti* unutar tablice 5-13 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih izgrađenih modela neovisno o udjelu ulaznih značajki i načinu uzorkovanja podataka. Drugim riječima, model će točno prepoznati (označiti) negativne primjerke, no iz mjere *osjetljivosti* je vidljivo kako to neće biti slučaj i s pozitivnim primjercima.

### 5.3.2. ATA 27

Transformacijom skupa podataka za izradu klasifikacijskih modela prve grupe, broj uzoraka u skupu  $D_{KMI-1}$  sustava ATA 27 za zrakoplove  $Ac_1$ ,  $Ac_2$ ,  $Ac_3$  i  $Ac_4$  je smanjen u odnosu na broj uzoraka u skupu  $D_{KMI}$  (slika 3-17). Zrakoplov  $Ac_1$  u skupu  $D_{KMI-1}$  bilježi 496 uzoraka ( $D_{KMI} = 960$  uzoraka),  $Ac_2$  1217 uzoraka ( $D_{KMI} = 2835$  uzoraka),  $Ac_3$  612 uzoraka ( $D_{KMI} = 1520$  uzoraka), a  $Ac_4$  196 uzoraka ( $D_{KMI} = 318$  uzoraka).

#### 5.3.2.1. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_1$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_1$  zabilježeno je 496 uzoraka, od čega je 473 uzoraka pripalo klasi 0, dok je preostalih 23 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 347 uzoraka, od kojih je 16 svrstano u klasu 1, a 331 u klasu 0. U skupu za verifikaciju (30%) svrstano je 149 uzoraka, od kojih je 142 pripalo klasi 0, a 7 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-14.

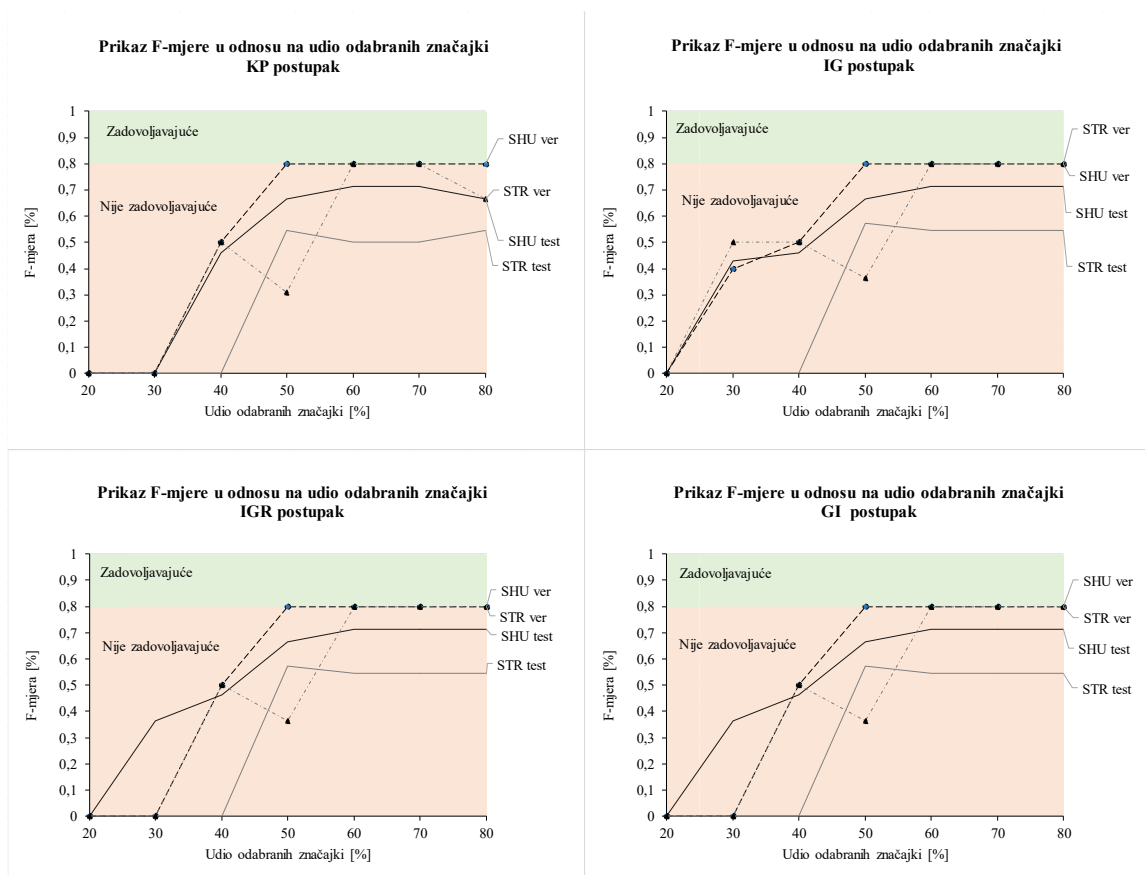
Tablica 5-14. ATA 27 zrakoplov  $Ac_1$  - vrijednosti *F*-mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka

	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,7692	<b>0,8000</b>	0,6000	0,6667
<b>Osjetljivost*</b>	0,6250	<b>0,6667</b>	0,4286	0,5000
<b>Specifičnost**</b>	1,0000	<b>1,0000</b>	1,0000	1,0000

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-14 je vidljivo kako klasifikacijski model izgrađen slučajnim uzorkovanjem podataka (SHU) ima bolje vrijednosti *F*-mjere i *osjetljivosti* u odnosu na model izgrađen stratificiranim uzorkovanjem podataka (STR) te je taj model uzet kao reprezentativan. Na slici 5-10 prikazane su vrijednosti *F*-mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano rang značajki  $R_{KMLi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ . Prag za *F*-mjeru je postavljen na 0,8 (80%). Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki od novoizgrađenih modela su prikazane u tablicama 5-15 i 5-16.



Slika 5-10. ATA 27 – prikaz vrijednosti *F*-mjere za  $Ac_1$

Iz slike 5-10 vidljivo je kako smanjenjem udjela ulaznih značajki neovisno o filterskom postupku i načinu uzorkovanja podataka, vrijednosti *F*-mjere neće prijeći zadani prag. Ista vrijednost mjere je postignuta kada su modeli izgrađeni slučajnim uzorkovanjem podataka, a udio ulaznih značajki smanjen na 50%, 60%, 70% i 80% neovisno o filterskom postupku. Također, iste vrijednosti *F*-mjere su postignute za slučaj kada su modeli izgrađeni stratificiranim uzorkovanjem podataka sa smanjenim udjelom značajki od 60%, 70% i 80% neovisno o filterskom postupku.

Tablica 5-15. ATA 27 – prikaz vrijednosti mjere osjetljivosti za  $Ac_1$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
30	0,0000	0,0000	0,0000	0,0000	0,2222	0,0000	0,0000	0,0000	0,3333	0,2500	0,0000	0,5000	0,2222	0,0000	0,0000	0,0000
40	0,3333	0,3333	0,0000	0,5000	0,3333	0,3333	0,0000	0,5000	0,3333	0,3333	0,0000	0,5000	0,3333	0,3333	0,0000	0,5000
50	0,5556	<b>0,6667</b>	0,7500	<b>1,0000</b>	0,5556	<b>0,6667</b>	0,7500	<b>1,0000</b>	0,5556	<b>0,6667</b>	0,7500	<b>1,0000</b>	0,5556	<b>0,6667</b>	0,7500	<b>1,0000</b>
60	0,6250	<b>0,6667</b>	0,3750	<b>1,0000</b>	0,6250	<b>0,6667</b>	0,3750	<b>1,0000</b>	0,6250	<b>0,6667</b>	0,3750	<b>1,0000</b>	0,6250	<b>0,6667</b>	0,3750	<b>1,0000</b>
70	0,6250	<b>0,6667</b>	0,3750	<b>1,0000</b>	0,6250	<b>0,6667</b>	0,3750	<b>1,0000</b>	0,6250	<b>0,6667</b>	0,3750	<b>1,0000</b>	0,6250	<b>0,6667</b>	0,3750	<b>1,0000</b>
80	0,5000	<b>0,6667</b>	0,3750	0,5000	0,6250	<b>0,6667</b>	0,3750	<b>1,0000</b>	0,6250	<b>0,6667</b>	0,3750	<b>1,0000</b>	0,6250	<b>0,6667</b>	0,3750	<b>1,0000</b>

Tablica 5-16. ATA 27 – prikaz vrijednosti mjere specifičnosti za  $Ac_1$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,9707	1,0000	0,9817	0,9756	1,0000	1,0000	0,9909	1,0000	0,9818	1,0000	0,9806	0,9750	0,9707	1,0000	0,9817	0,9750
30	0,9909	1,0000	0,9800	0,9756	1,0000	1,0000	0,9792	0,9750	0,9809	1,0000	0,9792	0,9750	1,0000	1,0000	0,9792	0,9750
40	0,9900	1,0000	0,9675	0,9750	0,9900	1,0000	0,9678	0,9744	0,9900	1,0000	0,9678	0,9744	0,9889	1,0000	0,9678	0,9744
50	0,9909	1,0000	0,9153	0,7568	0,9909	1,0000	0,9244	0,8056	0,9909	1,0000	0,9244	0,8205	0,9909	1,0000	0,9244	0,8205
60	0,9909	1,0000	0,9900	0,9714	0,9909	1,0000	1,0000	0,9714	0,9909	1,0000	1,0000	0,9714	0,9909	1,0000	1,0000	0,9714
70	0,9909	1,0000	0,9900	0,9706	0,9909	1,0000	1,0000	0,9706	0,9909	1,0000	1,0000	0,9706	0,9909	1,0000	1,0000	0,9706
80	1,0000	1,0000	1,0000	1,0000	0,9909	1,0000	1,0000	0,9706	0,9909	1,0000	1,0000	0,9706	0,9909	1,0000	1,0000	0,9706

Za vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-15 moguće je povući iste zaključke kao i za *F-mjeru*. Vidljivo je kako smanjenjem udjela ulaznih značajki na 50%, 60%, 70% i 80%, bez obzira na filterski postupak i način uzorkovanja podataka, novoizgrađeni modeli postižu jednake i bolje vrijednosti mjere *osjetljivosti* kao i model izgrađenim s ukupnim brojem ulaznih značajki (tablica 5-14).

Vrijednosti mjere *specifičnosti* unutar tablice 5-16 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih izgrađenih modela neovisno o udjelu ulaznih značajki i načinu uzorkovanja podataka.

### 5.3.2.2. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_2$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_2$  zabilježeno je 1217 uzoraka, od čega je 1185 uzoraka pripalo klasi 0, dok je preostalih 32 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 852 uzoraka, od kojih je 22 svrstano u klasu 1, a 830 u klasu 0. U skupu za verifikaciju (30%) svrstano je 365 uzoraka, od kojih je 355 pripalo klasi 0, a 10 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-17.

Tablica 5-17. ATA 27 zrakoplov  $Ac_2$  - vrijednosti *F*-mjere, *osjetljivosti* i *specifičnosti* za SHU i STR uzorkovanje podataka

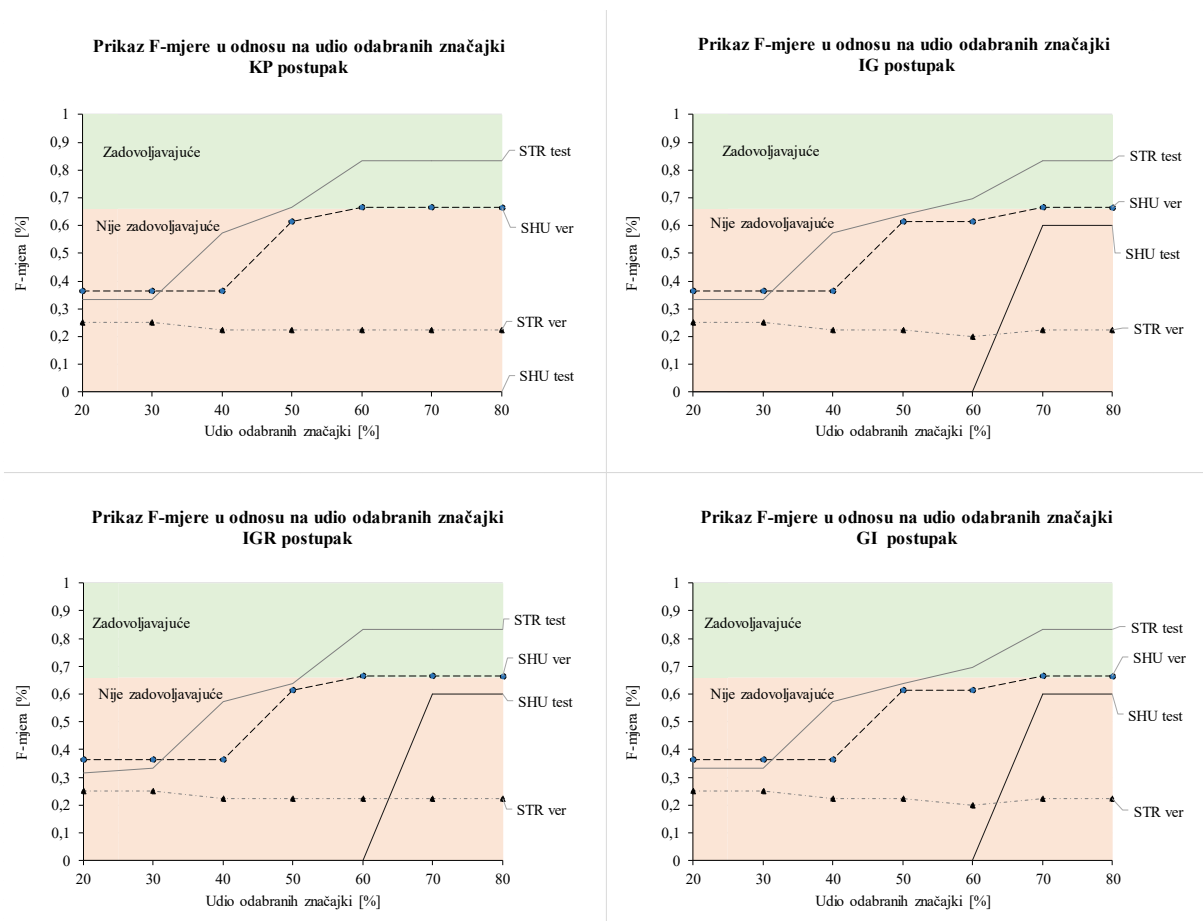
	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,6000	<b>0,6667</b>	0,8333	0,2222
<b>Osjetljivost*</b>	0,5000	<b>0,6250</b>	0,8500	0,1429
<b>Specifičnost**</b>	0,9913	<b>0,9792</b>	0,9865	0,9897

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-17 je vidljivo kako klasifikacijski model izgrađen slučajnim uzorkovanjem podataka (SHU) ima bolje vrijednosti statističkih mjera u odnosu na model izgrađen stratificiranim uzorkovanjem podataka (STR) te je taj model uzet kao reprezentativan. Na slici 5-11 prikazane su vrijednosti *F*-mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano rang značajki  $R_{KMIi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ . Prag za *F-mjeru* je postavljen na 0,66 (66%). Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki od novoizgrađenih modela su prikazane u tablicama 5-18 i 5-19.





Slika 5-11. ATA 27 – prikaz vrijednosti  $F$ -mjere za  $Ac_2$

Iz slike 5-11 vidljivo je kako smanjenjem udjela ulaznih značajki neovisno o filterskom postupku, vrijednosti  $F$ -mjere neće prijeći zadani prag.

Ista vrijednost te mjere je postignuta kada je model izgrađen slučajnim uzorkovanjem podataka, a udio ulaznih značajki je smanjen na 70% i 80% primjenom IG, IGR i GI filterskog postupka.

Iako su iste vrijednosti  $F$ -mjere postignute i kada je model izgrađen slučajnim uzorkovanjem s udjelom značajki smanjenim na 60% temeljem KP i IGR postupka, vidljivo je kako su testni rezultati  $F$ -mjere loši na tim skupovima podataka te ti modeli nisu reprezentativni.

Tablica 5-18. ATA 27 – prikaz vrijednosti mjere osjetljivosti za  $Ac_2$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,0000	0,2222	0,1500	0,1429	0,0000	0,2222	0,1500	0,1429	0,0000	0,2222	0,1500	0,1429	0,0000	0,2222	0,1500	0,1429
30	0,0000	0,2222	0,2500	0,1429	0,0000	0,2222	0,1500	0,1429	0,0000	0,2222	0,2500	0,1429	0,0000	0,2222	0,1500	0,1429
40	0,0000	0,2222	0,4500	0,1429	0,0000	0,2222	0,4500	0,1429	0,0000	0,2222	0,4500	0,1429	0,0000	0,2222	0,4000	0,1429
50	0,0000	0,5000	0,5714	0,1429	0,0000	0,5000	0,5500	0,1429	0,0000	0,5000	0,5500	0,1429	0,0000	0,5000	0,5500	0,1429
60	0,0000	0,6250	0,8500	0,1429	0,0000	0,6250	0,8500	0,1429	0,0000	0,5000	0,6500	0,1429	0,0000	0,5000	0,6500	0,1429
70	0,0000	0,6250	0,8500	0,1429	0,5000	<b>0,6250</b>	0,8500	0,1429	0,5000	<b>0,6250</b>	0,8500	0,1429	0,5000	<b>0,6250</b>	0,8500	0,1429
80	0,0000	0,6250	0,8500	0,1429	0,5000	<b>0,6250</b>	0,8500	0,1429	0,5000	<b>0,6250</b>	0,8500	0,1429	0,5000	<b>0,6250</b>	0,8500	0,1429

Tablica 5-19. ATA 27 – prikaz vrijednosti mjere specifičnosti za  $Ac_2$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	1,0000	1,0000	1,0000	1,0000	0,9958	1,0000	0,9958	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
30	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9958	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
40	0,9958	1,0000	1,0000	0,9898	0,9918	1,0000	1,0000	0,9898	0,9918	1,0000	1,0000	0,9898	0,9958	1,0000	1,0000	0,9898
50	0,9871	0,9898	0,9915	0,9897	0,9915	0,9899	0,9915	0,9898	0,9915	0,9900	0,9915	0,9898	0,9915	0,9900	0,9915	0,9898
60	0,9915	0,9796	0,9870	0,9897	0,9915	0,9794	0,9865	0,9897	0,9915	0,9897	0,9909	0,9794	0,9915	0,9897	0,9909	0,9794
70	0,9915	0,9796	0,9870	0,9897	0,9913	0,9794	0,9865	0,9897	0,9913	0,9794	0,9865	0,9897	0,9913	0,9794	0,9865	0,9897
80	0,9915	0,9796	0,9870	0,9897	0,9913	0,9794	0,9865	0,9897	0,9913	0,9794	0,9865	0,9897	0,9913	0,9794	0,9865	0,9897

Za vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-18 moguće je povući iste zaključke kao i za *F-mjeru*. Vidljivo je kako smanjenjem udjela ulaznih značajki na 70% i 80% temeljem IGR, IG i GI filterskih postupaka i primjenom slučajnog uzorkovanja podataka, novoizgrađeni modeli postižu jednake vrijednosti mjere *osjetljivosti* kao i model izgrađenim s ukupnim brojem ulaznih značajki (tablica 5-52).

Vrijednosti mjere *specifičnosti* unutar tablice 5-19 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih izgrađenih modela neovisno o udjelu ulaznih značajki i načinu uzorkovanja podataka.

### 5.3.2.3. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_3$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_3$  zabilježeno je 612 uzoraka, od čega je 595 uzoraka pripalo klasi 0, dok je preostalih 17 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 428 uzoraka, od kojih je 14 svrstano u klasu 1, a 414 u klasu 0. U skupu za verifikaciju (30%) svrstano je 184 uzoraka, od kojih je 178 pripalo klasi 0, a 6 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-20.

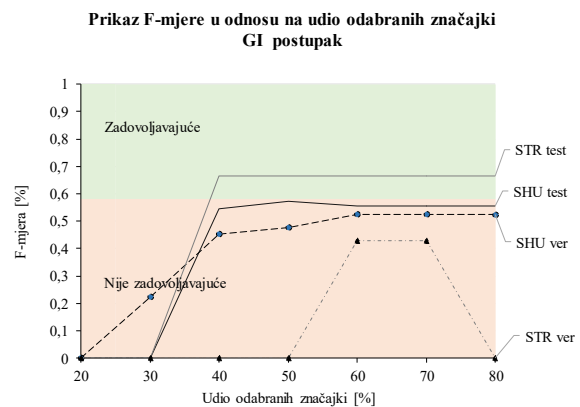
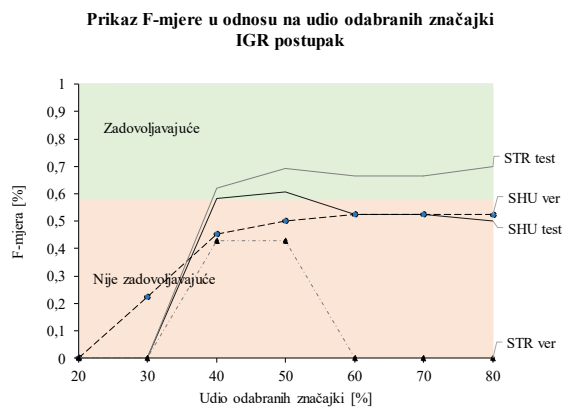
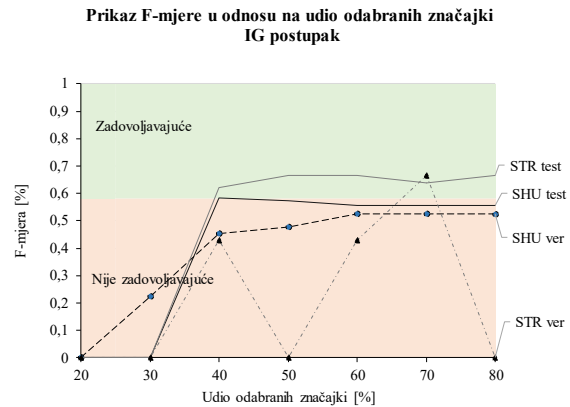
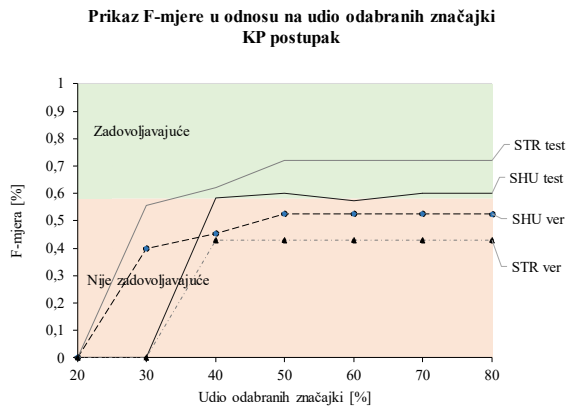
Tablica 5-20. ATA 27 zrakoplov  $Ac_3$  - vrijednosti *F*-mjere, *osjetljivosti* i *specifičnosti* za SHU i STR uzorkovanje podataka

	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,5556	<b>0,5882</b>	0,6957	0,5000
<b>Osjetljivost*</b>	0,5556	<b>1,0000</b>	0,6500	0,6000
<b>Specifičnost**</b>	0,9736	<b>0,8889</b>	0,9767	0,9322

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-20 je vidljivo kako klasifikacijski model izgrađen slučajnim uzorkovanjem podataka (SHU) ima bolje vrijednosti statističkih mjera u odnosu na model izgrađen stratificiranim uzorkovanjem podataka (STR) te je taj model uzet kao reprezentativan. Na slici 5-12 prikazane su vrijednosti *F*-mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano rang značajki  $R_{KMIi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ . Prag za *F-mjeru* je postavljen na 0,58 (58%). Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki od novoizgrađenih modela su prikazane u tablicama 5-21 i 5-22.



Slika 5-12. ATA 27 – prikaz vrijednosti  $F$ -mjere za  $Ac_3$

Iz slike 5-12 vidljivo je da je bolja vrijednost  $F$ -mjere postignuta jedino za slučaj kada je model izgrađen stratificiranim uzorkovanjem podataka sa smanjenim udjelom značajki na 70% primjenom IG filterskog postupka.

Tablica 5-21. ATA 27 – prikaz vrijednosti mjere osjetljivosti za  $Ac_3$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000
30	0,0000	0,4000	0,4000	0,0000	0,0000	0,2000	0,0000	0,0000	0,0000	0,2000	0,0000	0,0000	0,0000	0,2000	0,0000	0,0000
40	0,7000	<b>1,0000</b>	0,7500	0,6000	0,7000	<b>1,0000</b>	0,7500	0,6000	0,7000	<b>1,0000</b>	0,7500	0,6000	0,6667	<b>1,0000</b>	0,5500	0,0000
50	0,6667	<b>1,0000</b>	0,7500	0,6000	0,7000	<b>1,0000</b>	0,7000	0,6000	0,6667	<b>1,0000</b>	0,5500	0,0000	0,6667	<b>1,0000</b>	0,5500	0,0000
60	0,6667	<b>1,0000</b>	0,7500	0,6000	0,5000	<b>1,0000</b>	0,5500	0,0000	0,5556	<b>1,0000</b>	0,6500	0,6000	0,5556	<b>1,0000</b>	0,6500	0,6000
70	0,6667	<b>1,0000</b>	0,7500	0,6000	0,5000	<b>1,0000</b>	0,6000	0,0000	0,5556	<b>1,0000</b>	0,6500	0,6000	0,5556	<b>1,0000</b>	0,6500	0,6000
80	0,6667	<b>1,0000</b>	0,7500	0,6000	0,5000	<b>1,0000</b>	0,6500	0,0000	0,5556	<b>1,0000</b>	0,5500	0,0000	0,5556	<b>1,0000</b>	0,5500	0,0000

Tablica 5-22. ATA 27 – prikaz vrijednosti mjere specifičnosti za  $Ac_3$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
30	0,9871	0,9577	0,9933	1,0000	1,0000	0,9583	0,9938	0,9848	1,0000	0,9577	0,9933	0,9839	1,0000	0,9583	0,9938	0,9848
40	0,9541	0,8286	0,9511	0,9016	0,9536	0,8286	0,9511	0,9016	0,9536	0,8286	0,9511	0,9016	0,9536	0,8261	0,9870	1,0000
50	0,9670	0,8615	0,9719	0,9000	0,9603	0,8507	0,9674	0,9016	0,9603	0,8333	0,9866	1,0000	0,9603	0,8333	0,9866	1,0000
60	0,9603	0,8615	0,9719	0,9000	0,9736	0,8636	0,9866	1,0000	0,9736	0,8615	0,9703	0,9000	0,9736	0,8615	0,9703	0,9000
70	0,9670	0,8615	0,9703	0,9000	0,9736	0,8615	0,9867	1,0000	0,9736	0,8615	0,9824	0,9861	0,9736	0,8615	0,9703	0,9000
80	0,9670	0,8615	0,9703	0,9000	0,9670	0,8615	0,9923	1,0000	0,9736	0,8615	0,9866	1,0000	0,9736	0,8615	0,9866	1,0000

Za vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-21 vidljivo je kako smanjenjem udjela ulaznih značajki na 40%, 50%, 60%, 70% i 80%, bez obzira na filterski postupak uz primjenu slučajnog uzorkovanja podataka, novoizgrađeni modeli postižu jednake vrijednosti mjere *osjetljivosti* kao i model izgrađenim s ukupnim brojem ulaznih značajki (tablica 5-20). Dovodeći u korelaciju rezultate *F*-mjere i mjere *osjetljivosti* prikazanih u tablici je vidljivo da modeli izgrađeni s manjim brojem značajki točno klasificiraju pozitivne primjerke, no pri tome nisu precizni.

Vrijednosti mjere *specifičnosti* unutar tablice 5-22 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih izgrađenih modela neovisno o udjelu ulaznih značajki i načinu uzorkovanja podataka.

#### 5.3.2.4. Prikaz i evaluacija rezultata modela za zrakoplov *Ac4*

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova *Ac4* zabilježeno je 196 uzoraka, od čega je 125 uzoraka pripalo klasi 0, dok je preostalih 12 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 137 uzoraka, od kojih je 12 svrstano u klasu 1, a 125 u klasu 0. U skupu za verifikaciju (30%) svrstano je 59 uzoraka, od kojih je 53 pripalo klasi 0, a 6 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-23.

Tablica 5-23. *ATA 27 zrakoplov Ac4 - vrijednosti F-mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka*

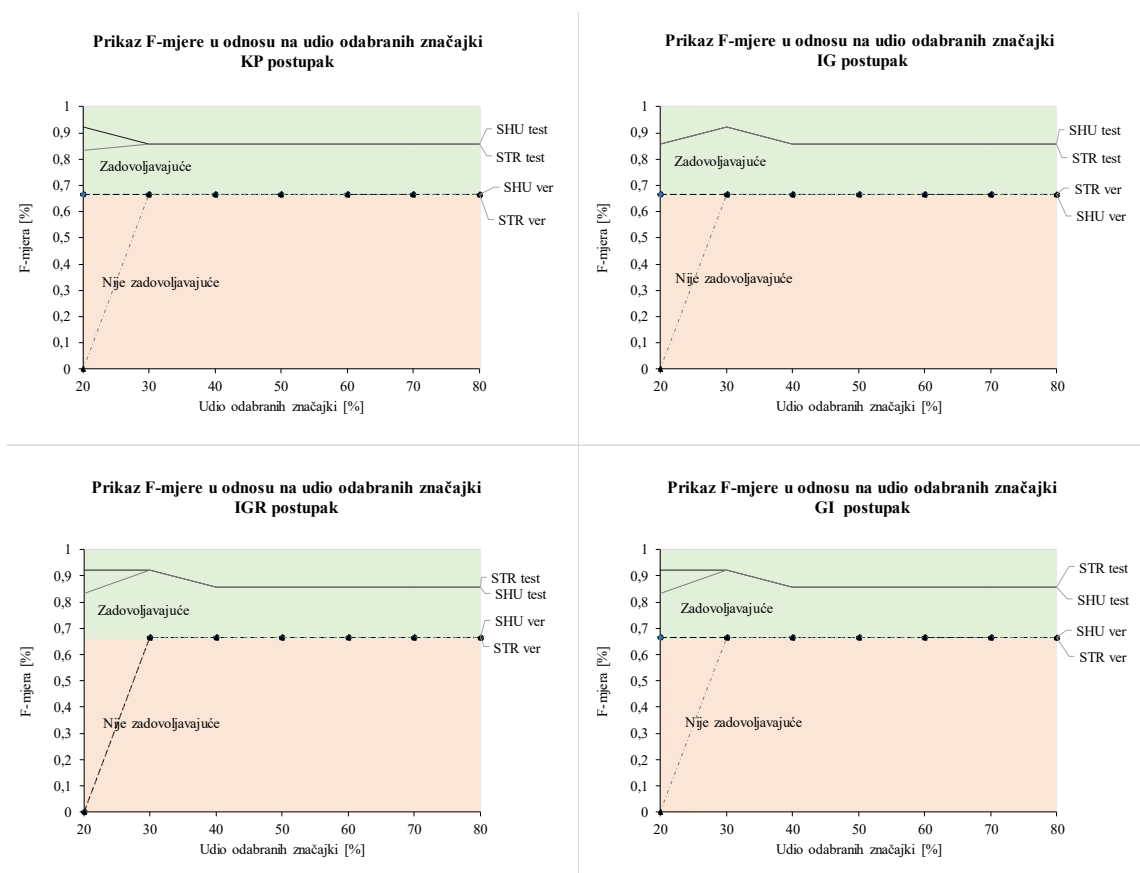
	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,8571	<b>0,6667</b>	0,8571	<b>0,6667</b>
<b>Osjetljivost*</b>	1,0000	<b>1,0000</b>	1,0000	<b>1,0000</b>
<b>Specifičnost**</b>	0,9167	<b>0,8750</b>	0,9300	<b>0,8750</b>

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-23 je vidljivo kako klasifikacijski model izgrađen stratificiranim uzorkovanjem podataka (STR) rezultira jednakim vrijednostima statističkih mjera kao i model izgrađen slučajnim uzorkovanjem podataka (SHU). Temeljem prikazanih vrijednosti statističkih mjera, smatra se kako su oba modela reprezentativna. Na slici 5-13 prikazane su vrijednosti *F*-mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem

učitanog ranga značajki  $R_{KMIi}$ , gdje je  $i = \text{KP, IG, IGR i GI}$ . Prag za  $F$ -mjeru je postavljen na 0,67 (67%). Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki od novoizgrađenih modela su prikazane u tablicama 5-24 i 5-25.



Slika 5-13. ATA 27 – prikaz vrijednosti  $F$ -mjere za  $Ac_4$

Iz slike 5-13 vidljivo je kako smanjenjem udjela ulaznih značajki neovisno o filterskom postupku i načinu uzorkovanja podataka, vrijednosti  $F$ -mjere neće prijeći zadani prag.

Ista vrijednost te mjere je postignuta već za slučaj kada je udio ulaznih značajki smanjen na 20% primjenom KP, IG i GI filterskog postupka i model izgrađen slučajnim uzorkovanjem podataka. Ta vrijednost ostaje nepromijenjena i za preostale udjele ulaznih značajki.

Također, iste vrijednosti  $F$ -mjere su postignute i za slučaj izgrađenih modela stratificiranim uzorkovanjem podataka uz smanjeni udio značajki od 30% do 80% neovisno o filterskom postupku.

Tablica 5-24. ATA 27 – prikaz vrijednosti mjere osjetljivosti za  $Ac_4$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	1,0000	1,0000	0,8333	0,0000	1,0000	0,0000	0,8333	0,0000	0,8571	1,0000	0,8571	0,0000	1,0000	1,0000	0,8333	0,0000
30	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
40	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
50	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
60	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
70	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
80	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tablica 5-25. ATA 27 – prikaz vrijednosti mjere specifičnosti za  $Ac_4$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,9667	0,9231	0,9750	0,9375	0,9667	0,9231	0,9750	0,9375	0,9667	0,9231	0,9750	0,9167	0,9667	0,9231	0,9750	0,9375
30	0,9500	0,9091	0,9167	0,9091	0,9500	0,9091	0,9500	0,9167	0,9500	0,9231	0,9667	0,9167	0,9500	0,9091	0,9500	0,9167
40	0,9333	0,9000	0,9167	0,9000	0,9333	0,9091	0,9167	0,9091	0,9333	0,9091	0,9167	0,9091	0,9333	0,9000	0,9167	0,9091
50	0,9333	0,8889	0,9130	0,8750	0,9333	0,9000	0,9130	0,8750	0,9333	0,8889	0,9167	0,8750	0,9333	0,8889	0,9167	0,8750
60	0,9333	0,8889	0,9130	0,8750	0,9333	0,9000	0,9091	0,8750	0,9333	0,8889	0,9091	0,8750	0,9333	0,8889	0,9091	0,8750
70	0,9333	0,8750	0,9048	0,8750	0,9333	0,8889	0,9091	0,8750	0,9333	0,8889	0,9091	0,8750	0,9333	0,8889	0,9091	0,8750
80	0,9333	0,8750	0,9048	0,8750	0,9333	0,8889	0,9048	0,8750	0,9333	0,8889	0,9333	0,8750	0,9333	0,8889	0,9333	0,8750



Za vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-24 moguće je povući iste zaključke kao i za *F-mjeru*. Gotovo svi modeli će rezultirati istim vrijednostima mjere osjetljivosti, neovisno o udjelu ulaznih značajki i filterskom postupku. Manje iznimke su vidljive kada je model izgrađen stratificiranim uzorkovanjem, a udio ulaznih značajki smanjen na 20% neovisno o filterskom postupku.

Vrijednosti mjere *specifičnosti* prikazanih u tablici 5-25 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih izgrađenih modela neovisno o udjelu ulaznih značajki i načinu uzorkovanja podataka.

### 5.3.3. ATA 32

Transformacijom skupa podataka za izradu klasifikacijskih modela prve grupe, broj uzoraka u skupu  $D_{KMI-1}$  sustava ATA 32 za zrakoplove  $Ac_1$ ,  $Ac_2$ ,  $Ac_3$  i  $Ac_4$  je smanjen u odnosu na broj uzoraka u skupu  $D_{KMI}$  (slika 3-17). Zrakoplov  $Ac_1$  u skupu  $D_{KMI-1}$  bilježi 145 uzoraka ( $D_{KMI} = 186$  uzoraka),  $Ac_2$  160 uzoraka ( $D_{KMI} = 199$  uzoraka),  $Ac_3$  152 uzoraka ( $D_{KMI} = 201$  uzorak), a  $Ac_4$  163 uzoraka ( $D_{KMI} = 183$  uzorka).

#### 5.3.3.1. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_1$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_1$  zabilježeno je 145 uzoraka, od čega je 112 uzoraka pripalo klasi 0, dok je preostalih 33 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 102 uzoraka, od kojih je 23 svrstano u klasu 1, a 78 u klasu 0. U skupu za verifikaciju (30%) svrstano je 43 uzoraka, od kojih je 34 pripalo klasi 0, a 10 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-26.

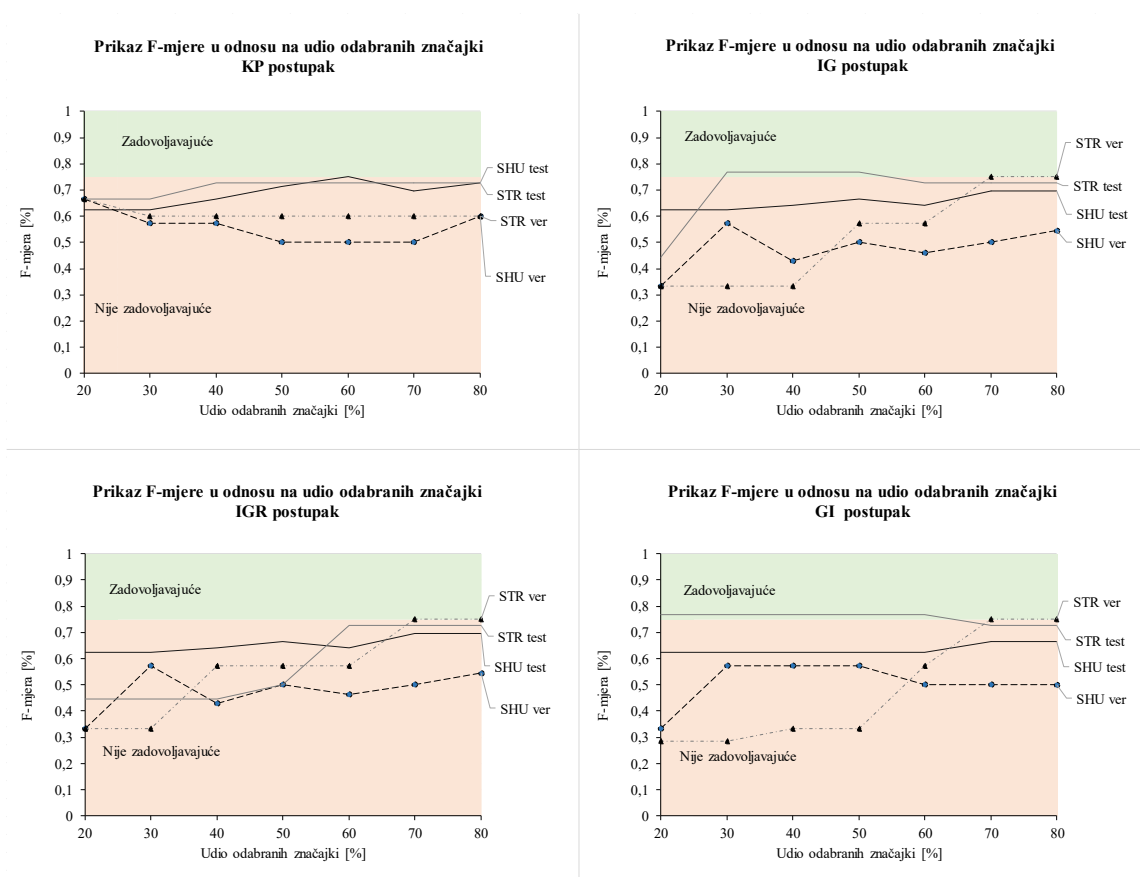
Tablica 5-26. ATA 32 zrakoplov  $Ac_1$  - vrijednosti *F*-mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka

	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,0000	0,8571	0,7273	<b>0,7500</b>
<b>Osjetljivost*</b>	0,0000	0,7500	0,6667	<b>0,7500</b>
<b>Specifičnost**</b>	0,8167	1,0000	0,9667	<b>0,9231</b>

\* mjera za klasu 1

\*\* mjera za klasu 0

Aalizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-26 je vidljivo kako klasifikacijski model izgrađen stratificiranim uzorkovanjem podataka (STR) rezultira nižim vrijednostima *F*-mjere i *specifičnosti* u odnosu na model izgrađen slučajnim uzorkovanjem podataka (SHU), no mjere tog modela na testnom skupu podataka su rezultirale boljim vrijednostima. Iz tog razloga je model izgrađen stratificiranim uzorkovanjem podataka uzet kao reprezentativan. Na slici 5-14 prikazane su vrijednosti *F*-mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano rang značajki  $R_{KMIi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ . Prag za *F*-mjeru je postavljen na 0,75 (75%). Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki od novoizgrađenih modela su prikazane u tablicama 5-27 i 5-28.



Slika 5-14. ATA 32 – prikaz vrijednosti *F*-mjere za  $Ac_1$

Iz slike 5-14 vidljivo je kako smanjenjem udjela ulaznih značajki neovisno o filterskom postupku, vrijednosti *F*-mjere neće prijeći zadani prag. Ista vrijednost te mjere je postignuta za slučaj kada je udio ulaznih značajki je smanjen na 70% i 80% primjenom IG, IGR i GI filterskog postupka i model izgrađen stratificiranim uzorkovanjem podataka.

Tablica 5-27. ATA 32 – prikaz vrijednosti mjere osjetljivosti za  $Ac_1$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,5000	0,6667	0,5000	<b>0,7500</b>	0,5000	0,3333	0,2857	0,2500	0,5000	0,3333	0,2857	0,2500	0,5000	0,3333	0,7143	0,2500
30	0,5000	0,6667	0,5000	<b>0,7500</b>	0,5000	0,6667	0,2857	0,2500	0,5000	0,6667	0,7143	0,2500	0,5556	0,6667	0,7143	0,2500
<b>40</b>	0,5556	0,6667	0,5714	<b>0,7500</b>	0,9000	<b>1,0000</b>	0,2857	0,5000	0,9000	<b>1,0000</b>	0,7143	0,2500	0,5556	0,6667	0,7143	0,2500
50	0,5556	0,6667	0,5714	<b>0,7500</b>	0,5556	0,6667	0,3333	0,5000	0,5556	0,6667	0,7143	0,5000	0,5556	0,6667	0,7143	0,2500
<b>60</b>	0,6667	0,6667	0,5714	<b>0,7500</b>	0,8889	<b>1,0000</b>	0,6667	0,5000	0,8889	<b>1,0000</b>	0,6667	0,5000	0,5556	0,6667	0,7143	0,5000
<b>70</b>	0,8889	<b>1,0000</b>	0,5714	<b>0,7500</b>	0,8889	<b>1,0000</b>	0,6667	<b>0,7500</b>	0,8889	<b>1,0000</b>	0,6667	<b>0,7500</b>	0,6250	0,6667	0,6667	<b>0,7500</b>
<b>80</b>	0,8889	<b>1,0000</b>	0,5714	<b>0,7500</b>	0,8889	<b>1,0000</b>	0,6667	<b>0,7500</b>	0,8889	<b>1,0000</b>	0,6667	<b>0,7500</b>	0,8889	<b>1,0000</b>	0,6667	<b>0,7500</b>

Tablica 5-28. ATA 32 – prikaz vrijednosti mjere specifičnosti za  $Ac_1$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,9667	0,9231	1,0000	0,8000	0,9667	0,8333	1,0000	0,9231	0,9667	0,8333	1,0000	0,9231	0,9667	0,8333	0,9667	0,8462
30	0,9667	0,8462	1,0000	0,7000	0,9667	0,8333	1,0000	0,9231	0,9667	0,8333	0,9667	0,9231	0,9200	0,8333	0,9667	0,8462
40	0,9667	0,8333	1,0000	0,7000	0,7000	0,3333	1,0000	0,9231	0,7000	0,3333	0,9667	0,9231	0,9167	0,8333	0,9667	0,9231
50	1,0000	0,7500	1,0000	0,7000	0,7000	0,7500	1,0000	0,9231	0,9667	0,7500	0,9667	0,9231	0,9333	0,8333	0,9667	0,9231
60	0,9500	0,7500	1,0000	0,7000	0,7167	0,4167	0,9667	0,9231	0,7167	0,4167	0,9667	0,9231	0,9333	0,7500	0,9667	0,9231
70	0,7833	0,5000	1,0000	0,7000	0,7833	0,5000	0,9667	0,9231	0,7833	0,5000	0,9667	0,9231	0,9333	0,7500	0,9667	0,9231
80	0,8167	0,6667	1,0000	0,7000	0,7833	0,5833	0,9667	0,9231	0,7833	0,5833	0,9667	0,9231	0,7000	0,5000	0,9667	0,9231

Za vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-27 vidljivo je kako smanjenjem udjela ulaznih značajki na bez obzira na filterski postupak uz primjenu stratificiranog uzorkovanja podataka, novoizgrađeni modeli će postići iste, ali ne i bolje vrijednosti mjere *osjetljivosti*. Bolje vrijednosti mjere se postižu kada je model izgrađen slučajnim uzorkovanjem podataka za udio ulaznih značajki označen u tablici.

Vrijednosti mjere *specifičnosti* unutar tablice 5-28 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih modela izgrađenih stratificiranim uzorkovanjem podataka primjenom IGR, IG i GI filterskih postupaka bez obzira na udio ulaznih značajki.

### 5.3.3.2. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_2$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_2$  zabilježeno je 160 uzoraka, od čega je 125 uzoraka pripalo klasi 0, dok je preostalih 35 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 112 uzoraka, od kojih je 25 svrstano u klasu 1, a 88 u klasu 0. U skupu za verifikaciju (30%) svrstano je 48 uzoraka, od kojih je 37 pripalo klasi 0, a 10 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-29.

Tablica 5-29. ATA 32 zrakoplov  $Ac_2$  - vrijednosti  $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka

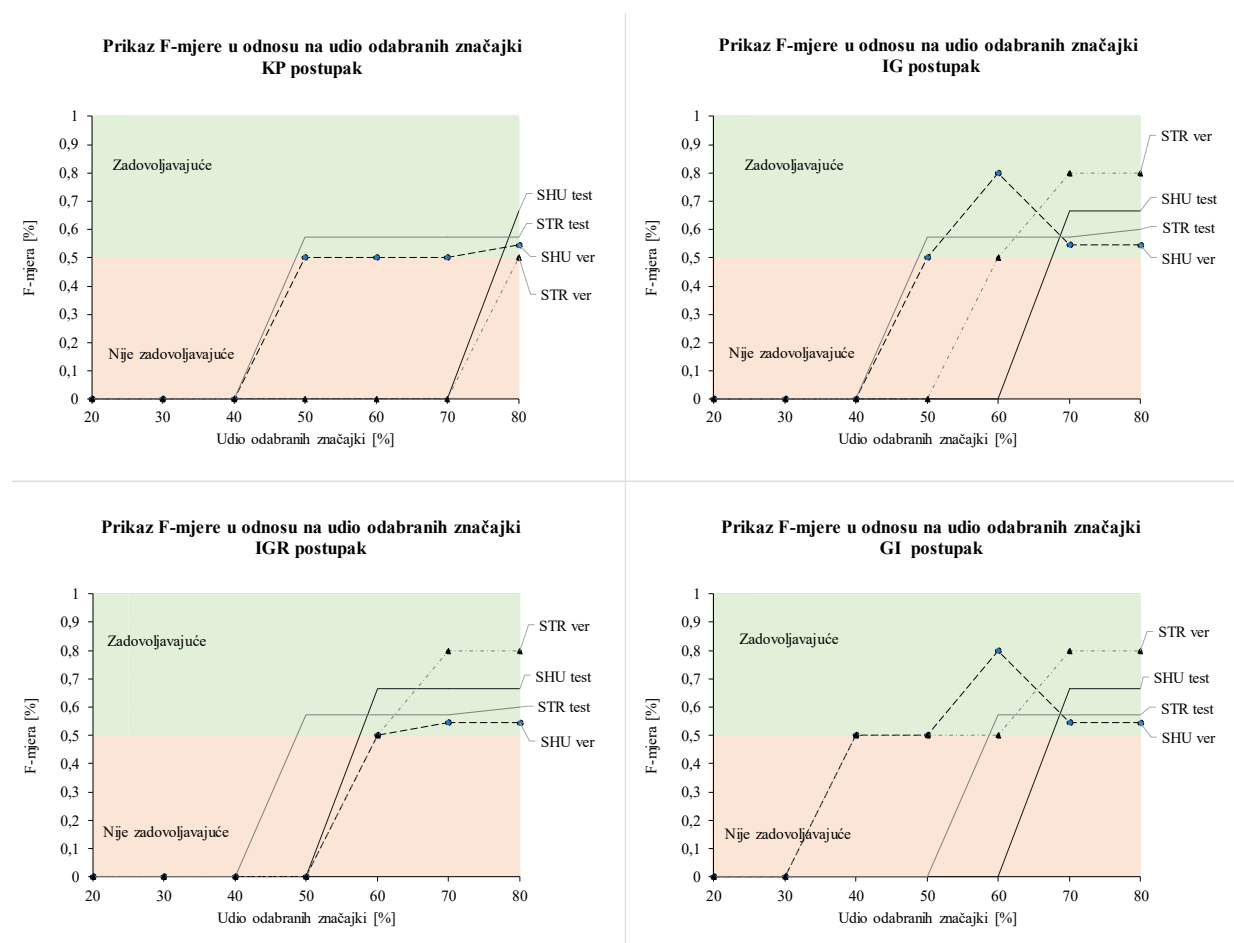
	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,6667	0,2222	0,5926	<b>0,5000</b>
<b>Osjetljivost*</b>	0,8182	0,5000	0,8889	<b>0,6667</b>
<b>Specifičnost**</b>	0,7750	0,6471	0,7083	<b>0,6000</b>

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-29 je vidljivo kako klasifikacijski model izgrađen stratificiranim uzorkovanjem podataka (STR) rezultira boljim vrijednostima statističkih mjera u odnosu na model izgrađen slučajnim uzorkovanjem podataka (SHU) te je taj model je uzet kao reprezentativan. Na slici 5-15 prikazane su vrijednosti  $F$ -mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano rang značajki  $R_{KMIi}$ , gdje je  $i = KP, IG$ ,

IGR i GI. Prag za  $F$ -mjeru je postavljen na 0,5 (50%). Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki novoizgrađeni model su prikazane u tablicama 5-30 i 5-31.



Slika 5-15. ATA 32 – prikaz vrijednosti  $F$ -mjere za  $Ac_2$

Iz slike 5-15 vidljivo je kako modeli izgrađeni stratificiranim uzorkovanjem podataka sa smanjenim udjelom značajki na 70% i 80%, uz primjenu IG, IGR i GI filterskog postupka, postižu bolje vrijednosti  $F$ -mjere u odnosu model izgrađen s ukupnim brojem značajki.

Bolje vrijednosti  $F$ -mjere također postižu modeli izgrađeni slučajnim uzorkovanjem kada je udio ulaznih značajki smanjen na 60% primjenom IG i IGR filterskih postupaka.

Tablica 5-30. ATA 32 – prikaz vrijednosti mjere osjetljivosti za  $Ac_2$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	STR test	STR ver
20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
30	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
40	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,3333	0,0000	0,3333
50	0,0000	0,3333	0,4000	0,0000	0,0000	0,0000	0,4000	0,0000	0,0000	0,3333	0,4000	0,0000	0,0000	0,3333	0,0000	0,3333
60	0,0000	0,3333	0,4000	0,0000	1,0000	<b>1,0000</b>	0,4000	0,3333	0,0000	0,6667	0,4000	0,3333	0,0000	0,6667	0,4000	0,3333
70	0,0000	0,3333	0,4000	0,0000	1,0000	<b>1,0000</b>	0,4000	<b>0,6667</b>	1,0000	<b>1,0000</b>	0,4000	<b>0,6667</b>	1,0000	<b>1,0000</b>	0,4000	<b>0,6667</b>
80	1,0000	<b>1,0000</b>	0,4000	0,3333	1,0000	<b>1,0000</b>	0,6000	<b>0,6667</b>	1,0000	<b>1,0000</b>	0,6000	<b>0,6667</b>	1,0000	<b>1,0000</b>	0,4000	<b>0,6667</b>

Tablica 5-31. ATA 32 – prikaz vrijednosti mjere specifičnosti za  $Ac_2$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	1,0000	1,0000	0,9750	1,0000	0,9000	1,0000	0,7000	1,0000	1,0000	1,0000	0,9750	1,0000	1,0000	1,0000	0,9750	1,0000
30	1,0000	1,0000	0,9750	1,0000	1,0000	1,0000	0,9750	1,0000	1,0000	1,0000	0,9750	1,0000	1,0000	1,0000	0,9750	1,0000
40	1,0000	1,0000	0,8800	1,0000	1,0000	1,0000	0,7933	1,0000	1,0000	1,0000	0,7933	1,0000	1,0000	1,0000	0,8683	1,0000
50	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9750	1,0000
60	1,0000	1,0000	1,0000	1,0000	0,8667	0,6250	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
70	1,0000	1,0000	1,0000	1,0000	0,8667	0,6875	1,0000	1,0000	0,8667	0,6875	1,0000	1,0000	0,8667	0,6875	1,0000	1,0000
80	0,8667	0,6875	1,0000	1,0000	0,8667	0,6875	0,9417	1,0000	0,8667	0,6875	0,9417	1,0000	0,8667	0,6875	1,0000	1,0000

Iz tablice 5-30 je vidljivo kako modeli izgrađeni stratificiranim uzorkovanjem podataka kada je udio značajki smanjen na 70% i 80% temeljem IGR, IG i GI filterskog postupka, postižu jednake, ali ne i bolje, vrijednosti mjere *osjetljivosti*. Za razliku od toga, dio modela izgrađenih slučajnim uzorkovanjem podataka postiže maksimalne vrijednosti mjere *osjetljivosti*, što je i naznačeno u tablici.

Vrijednosti mjere *specifičnosti* unutar tablice 5-31 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih izgrađenih modela neovisno o udjelu ulaznih značajki i načinu uzorkovanja podataka.

### 5.3.3.3. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_3$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_3$  zabilježeno je 152 uzoraka, od čega je 122 uzoraka pripalo klasi 0, dok je preostalih 30 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 106 uzoraka, od kojih je 20 svrstano u klasu 1, a 86 u klasu 0. U skupu za verifikaciju (30%) svrstano je 46 uzoraka, od kojih je 36 pripalo klasi 0, a 10 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-32.

Tablica 5-32. ATA 32 zrakoplov  $Ac_3$  - vrijednosti  $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka

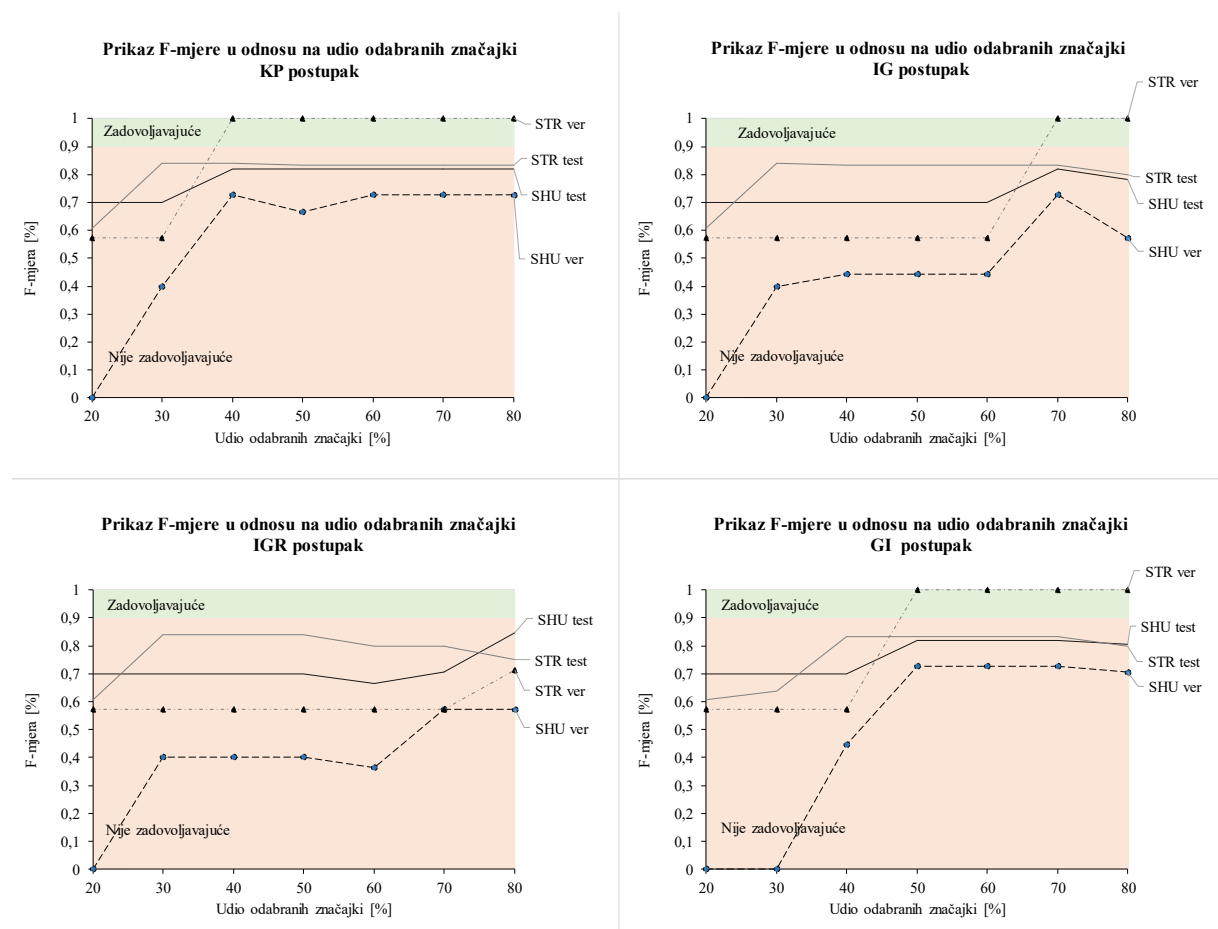
	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,9067	0,5714	0,7407	<b>1,0000</b>
<b>Osjetljivost*</b>	0,9500	0,6667	0,7500	<b>1,0000</b>
<b>Specifičnost**</b>	0,9300	0,8000	0,9067	<b>1,0000</b>

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-32 je vidljivo kako klasifikacijski model izgrađen stratificiranim uzorkovanjem podataka (STR) rezultira boljim vrijednostima statističkih mjera u odnosu na model izgrađen slučajnim uzorkovanjem podataka (SHU) te je taj model je uzet kao reprezentativan. Na slici 5-16 prikazane su vrijednosti  $F$ -mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano rang značajki  $R_{KMIi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ . Iako je vrijednost  $F$ -mjere dobivena modelom izgrađenim s ukupnim brojem značajki iznosila 1,0 (100%), prag te mjere je postavljen na 0,9 (90%) kako bi se ispitalo da li postoji

model koji je brži, tj. ima manji broj ulaznih značajki, a da isto postiže maksimalne vrijednosti *F*-mjere. Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki novoizgrađeni model su prikazane u tablicama 5-33 i 5-34.



Slika 5-16. ATA 32 – prikaz vrijednosti *F*-mjere za  $Ac_3$

Iz slike 5-16 je vidljivo kako je maksimalna vrijednost *F*-mjere postignuta već za slučaj kada je model izgrađen stratificiranim uzorkovanjem podataka sa smanjenim udjelom značajki na 40% primjenom KP filterskog postupka. Ta vrijednost ostaje nepromijenjena i za udjele značajki veće od 40%, kada se primjenjuje isti filterski postupak i stratificirano uzorkovanje podataka.

Također, modeli izgrađeni stratificiranim uzorkovanjem podataka sa smanjenim udjelom ulaznih značajki od 50% do 80% primjenom GI filterskog postupka te 70% i 80% primjenom IG postupka rezultiraju maksimalnim vrijednostima *F*-mjere.



Tablica 5-33. ATA 32 – prikaz vrijednosti mjere osjetljivosti za  $Ac_3$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,5500	0,0000	0,5000	0,4000	0,5500	0,0000	0,5000	0,4000	0,5500	0,0000	0,5000	0,4000	0,5500	0,0000	0,5000	0,4000
30	0,5500	0,3333	0,8500	0,4000	0,5500	0,3333	0,8500	0,4000	0,5500	0,3333	0,8500	0,4000	0,6000	0,0000	0,5500	0,4000
40	0,7000	0,6667	0,8500	<b>1,0000</b>	0,5500	0,3333	0,8500	0,4000	0,5000	0,3333	0,8500	0,4000	0,5000	0,3333	0,8500	0,4000
50	0,8000	0,6667	0,8500	<b>1,0000</b>	0,5500	0,3333	0,8500	0,4000	0,6000	0,3333	0,8500	0,4000	0,7500	0,6667	0,8500	<b>1,0000</b>
60	0,7500	0,6667	0,8500	<b>1,0000</b>	0,5000	0,3333	0,7500	0,4000	0,6000	0,3333	0,8500	0,4000	0,7500	0,6667	0,8500	<b>1,0000</b>
70	0,7500	0,6667	0,8500	<b>1,0000</b>	1,0000	<b>1,0000</b>	0,7500	0,4000	0,7500	0,6667	0,8500	<b>1,0000</b>	0,7500	0,6667	0,8500	<b>1,0000</b>
80	0,7500	0,6667	0,8500	<b>1,0000</b>	1,0000	0,6667	0,9500	<b>1,0000</b>	0,7500	0,6667	0,7500	<b>1,0000</b>	1,0000	<b>1,0000</b>	0,7500	<b>1,0000</b>

Tablica 5-34. ATA 32 – prikaz vrijednosti mjere specifičnosti za  $Ac_3$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,9800	0,9000	0,9400	1,0000	0,9800	0,9000	0,9400	1,0000	0,9800	0,9000	0,9400	1,0000	0,9800	0,9000	0,9400	1,0000
30	0,9800	0,9000	0,9400	1,0000	0,9800	0,9000	0,9400	1,0000	0,9800	0,9000	0,9400	1,0000	0,9800	0,9500	0,9600	1,0000
40	0,9800	0,9500	0,9400	1,0000	0,9800	0,9000	0,9400	1,0000	0,9800	0,9500	0,9350	1,0000	0,9800	0,9500	0,9350	1,0000
50	0,9800	0,9000	0,9350	1,0000	0,9800	0,9000	0,9400	1,0000	0,9800	0,9500	0,9350	1,0000	0,9800	0,9500	0,9350	1,0000
60	0,9800	0,9500	0,9350	1,0000	0,9550	0,8500	0,9600	1,0000	0,9800	0,9500	0,9350	1,0000	0,9800	0,9500	0,9350	1,0000
70	0,9800	0,9500	0,9350	1,0000	0,6550	0,5500	0,9600	1,0000	0,9800	0,9500	0,9350	1,0000	0,9800	0,9500	0,9350	1,0000
80	0,9800	0,9500	0,9350	1,0000	0,8350	0,8000	0,7700	0,8095	0,9600	0,8000	0,9600	1,0000	0,8100	0,7500	0,9600	1,0000

Iz tablice 5-33 vidljivo je kako smanjenjem udjela ulaznih značajki već na 40% primjenom KP filterskog postupka, model izgrađen stratificiranjem uzorkovanjem podataka postiže maksimalne vrijednosti mjere *osjetljivosti*. Te vrijednosti, kod primjene istog filterskog postupka i načina uzorkovanja podataka, ostaju nepromijenjene i kada je udio ulaznih značajki veći od 40%. Maksimalna vrijednost mjere *osjetljivosti* je također postignuta kod svih modela koji su izgrađeni stratificiranim uzorkovanjem, a udio ulaznih značajki je smanjen na 80% neovisno o filterskom postupku.

Vrijednosti mjere *specifičnosti* unutar tablice 5-34 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih izgrađenih modela neovisno o udjelu ulaznih značajki i načinu uzorkovanja podataka.

#### 5.3.3.4. *Prikaz i evaluacija rezultata modela za zrakoplov Ac4*

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_4$  zabilježeno je 163 uzoraka, od čega je 126 uzoraka pripalo klasi 0, dok je preostalih 37 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 114 uzoraka, od kojih je 26 svrstano u klasu 1, a 88 u klasu 0. U skupu za verifikaciju (30%) svrstano je 49 uzoraka, od kojih je 38 pripalo klasi 0, a 11 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-35.

Tablica 5-35. *ATA 32 zrakoplov Ac4 - vrijednosti F-mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka*

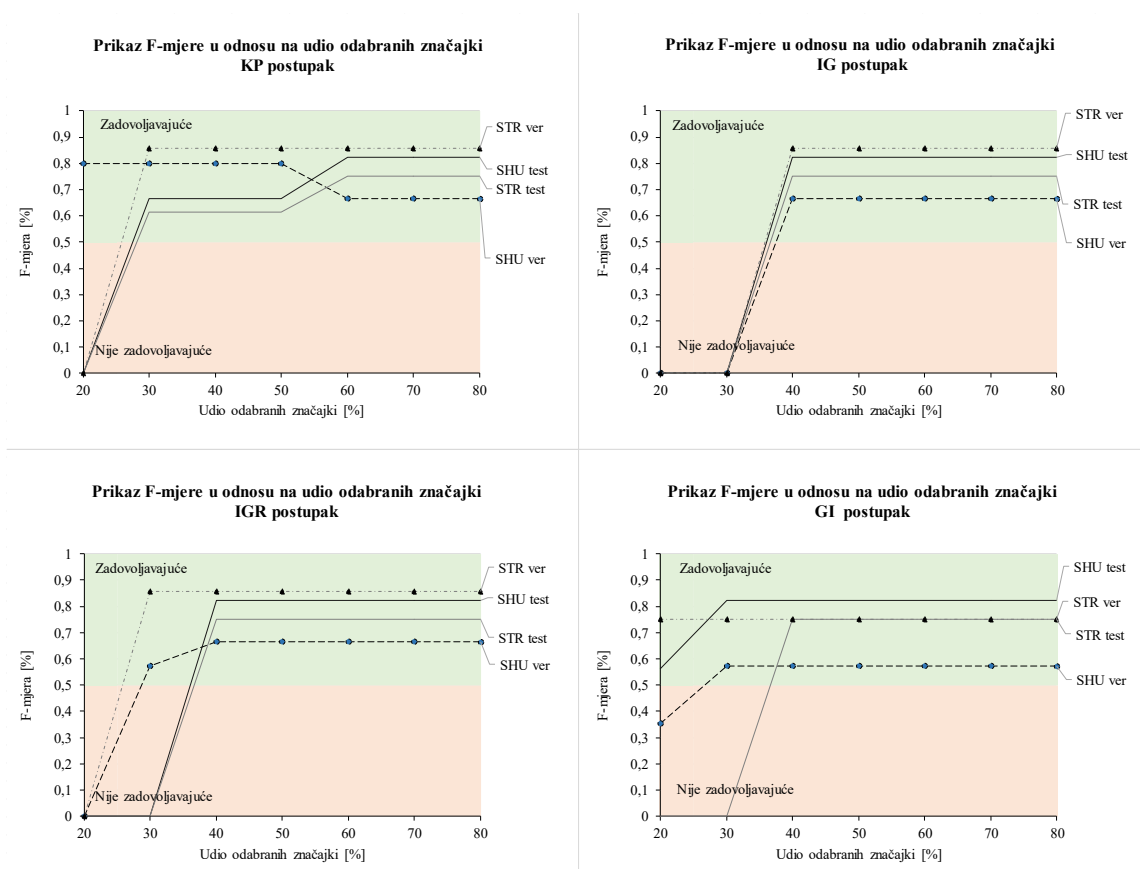
	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,8235	<b>0,5000</b>	0,7273	0,0000
<b>Osjetljivost*</b>	0,7778	<b>0,3333</b>	0,6667	0,0000
<b>Specifičnost**</b>	1,0000	<b>1,0000</b>	0,9750	1,0000

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-35 je vidljivo kako klasifikacijski model izgrađen slučajnim uzorkovanjem podataka (SHU) ima bolje vrijednosti statističkih mjera u odnosu na model izgrađen stratificiranim uzorkovanjem podataka (STR) te je taj model uzet kao reprezentativan. Na slici 5-17 prikazane su vrijednosti *F-mjere* dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano rang ranga značajki  $R_{KMIi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ .

Prag za  $F$ -mjeru je postavljen na 0,5 (50%). Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki od novoizgrađenih modela su prikazane u tablicama 5-36 i 5-37.



Slika 5-17. ATA 32 – prikaz vrijednosti  $F$ -mjere za  $Ac_4$

Iz slike 5-16 je vidljivo kako svi modeli sa smanjenim udjelom značajki od 40% do 80%, bez obzira na primijenjeni filterski postupak i način uzorkovanja podataka, postižu bolje vrijednosti  $F$ -mjere od modela koji je izgrađen s ukupnim brojem ulaznih značajki. Također je vidljivo kako smanjenjem udjela ulaznih značajki već na 30% primjenom KP postupka, model izgrađen slučajnim uzorkovanjem podataka daje bolju vrijednost  $F$ -mjere. Isti zaključci vrijede i za slučaj kada su ulazne značajke filtrirane GI postupkom u istom udjelu, a model izgrađen stratificiranjem uzorkovanjem podataka.

Tablica 5-36. ATA 32 – prikaz vrijednosti mjere osjetljivosti za  $Ac_4$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,0000	0,6667	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	<b>1,0000</b>	0,0000	0,7500
30	0,5556	<b>0,6667</b>	0,5000	<b>0,7500</b>	0,0000	0,6667	0,0000	0,7500	0,0000	0,0000	0,0000	0,0000	0,7778	<b>0,6667</b>	0,0000	0,7500
40	0,5556	<b>0,6667</b>	0,5000	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>
50	0,5556	<b>0,6667</b>	0,5000	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>
60	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>
70	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>
80	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>	0,7778	<b>0,6667</b>	0,7500	<b>0,7500</b>

Tablica 5-37. ATA 32 – prikaz vrijednosti mjere specifičnosti za  $Ac_4$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,8667	1,0000	0,9333	1,0000	1,0000	0,9231	0,9333	1,0000	0,9333	1,0000	0,9167	1,0000	0,4667	0,1538	1,0000	0,9167
30	0,9500	1,0000	0,9667	1,0000	0,9000	0,8462	0,9000	1,0000	0,9333	0,9231	0,8500	1,0000	0,9667	0,8462	0,8333	0,9167
40	0,9500	1,0000	0,9667	1,0000	0,9667	0,9231	0,9333	1,0000	0,9667	0,9231	0,9333	1,0000	0,9667	0,8462	0,9333	0,9167
50	0,9500	1,0000	0,9667	1,0000	0,9667	0,9231	0,9333	1,0000	0,9667	0,9231	0,9333	1,0000	0,9667	0,8462	0,9333	0,9167
60	0,9500	0,9231	0,9333	1,0000	0,9667	0,9231	0,9333	1,0000	0,9667	0,9231	0,9333	1,0000	0,9667	0,8462	0,9333	0,9167
70	0,9500	0,9231	0,9333	1,0000	0,9667	0,9231	0,9333	1,0000	0,9667	0,9231	0,9333	1,0000	0,9667	0,8462	0,9333	0,9167
80	0,9500	0,9231	0,9333	1,0000	0,9500	0,9231	0,9333	1,0000	0,9500	0,9231	0,9333	1,0000	0,9667	0,8462	0,9333	0,9167

Za vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-36 vidljivo je kako smanjenjem udjela ulaznih značajki na 40%, 50%, 60%, 70% i 80%, bez obzira na filterski postupak i način uzorkovanja podataka, novoizgrađeni modeli postižu jednake vrijednosti mjere *osjetljivosti* kao i model izgrađenim s ukupnim brojem ulaznih značajki. Također je vidljivo kako smanjenjem udjela ulaznih značajki već na 30% primjenom KP i GI postupka, modeli izgrađeni slučajnim uzorkovanjem podataka daje bolju vrijednost mjere *osjetljivosti*.

Vrijednosti mjere *specifičnosti* unutar tablice 5-37 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih izgrađenih modela neovisno o udjelu ulaznih značajki i načinu uzorkovanja podataka.

#### 5.3.4. ATA 34

Transformacijom skupa podataka za izradu klasifikacijskih modela prve grupe, broj uzoraka u skupu  $D_{KMI-1}$  sustava ATA 34 za zrakoplove  $Ac_1$ ,  $Ac_2$ ,  $Ac_3$  i  $Ac_4$  je smanjen u odnosu na broj uzoraka u skupu  $D_{KMI}$  (slika 3-17). Zrakoplov  $Ac_1$  u skupu  $D_{KMI-1}$  bilježi 370 uzoraka ( $D_{KMI} = 497$  uzoraka),  $Ac_2$  290 uzoraka ( $D_{KMI} = 343$  uzoraka),  $Ac_3$  570 uzoraka ( $D_{KMI} = 864$  uzoraka), a  $Ac_4$  622 uzoraka ( $D_{KMI} = 800$  uzoraka).

##### 5.3.4.1. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_1$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_1$  zabilježeno je 370 uzoraka, od čega je 351 uzoraka pripalo klasi 0, dok je preostalih 19 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 259 uzoraka, od kojih je 13 svrstano u klasu 1, a 246 u klasu 0. U skupu za verifikaciju (30%) svrstano je 111 uzoraka, od kojih je 105 pripalo klasi 0, a 6 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-38.

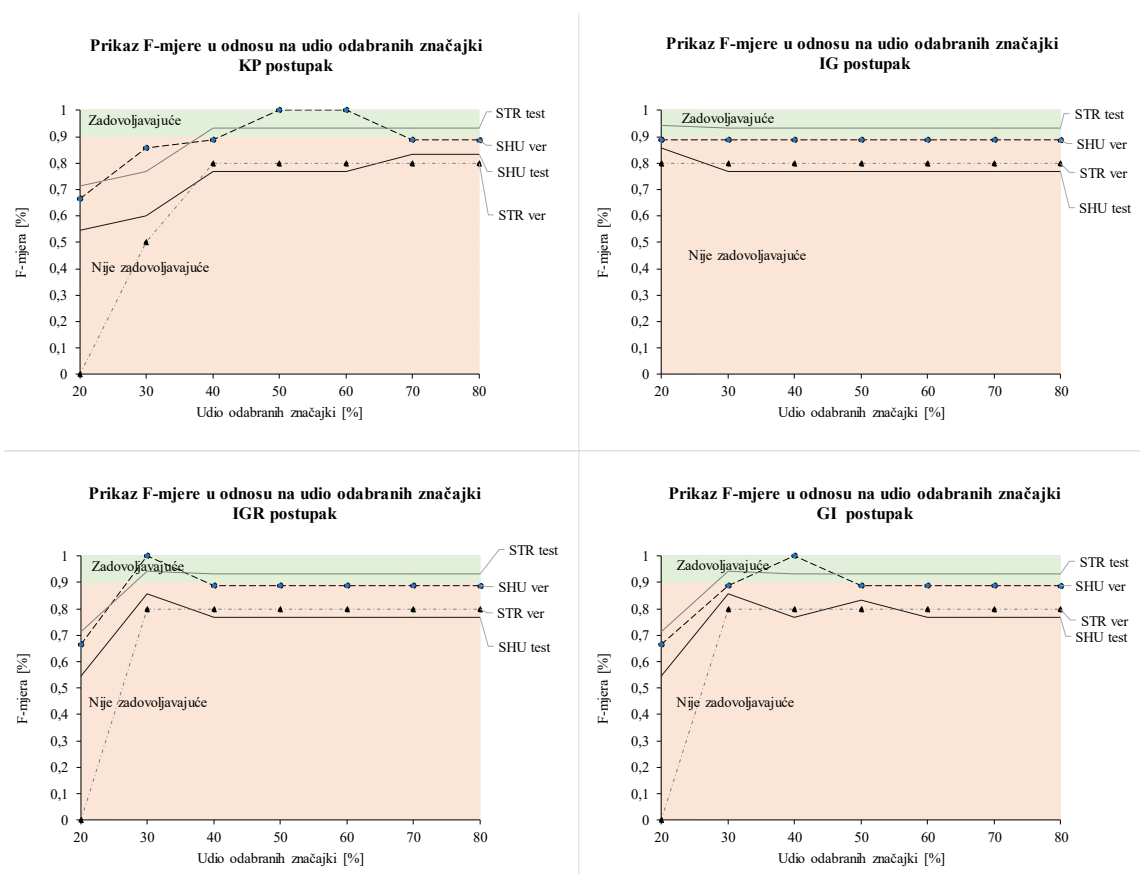
Tablica 5-38. ATA 34 zrakoplov  $Ac_1$  - vrijednosti *F*-mjere, *osjetljivosti* i *specifičnosti* za SHU i STR uzorkovanje podataka

	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,7692	<b>1,0000</b>	0,9333	0,8000
<b>Osjetljivost*</b>	0,7143	<b>1,0000</b>	1,0000	0,6667
<b>Specifičnost**</b>	0,9875	<b>1,0000</b>	0,9875	1,0000

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-38 je vidljivo kako klasifikacijski model izgrađen slučajnim uzorkovanjem podataka (SHU) ima bolje vrijednosti statističkih mjera u odnosu na model izgrađen stratificiranim uzorkovanjem podataka (STR) te je taj model uzet kao reprezentativan. Na slici 5-18 prikazane su vrijednosti  $F$ -mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano rang značajki  $R_{KMLi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ . Iako je vrijednost  $F$ -mjere dobivena modelom izgrađenim s ukupnim brojem značajki iznosila 1,0 (100%), prag te mjere je postavljen na 0,9 (90%) kako bi se ispitalo da li postoji model koji je brži, tj. ima manji broj ulaznih značajki, a da isto postiže visoke vrijednosti  $F$ -mjere. Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki novoizgrađeni model su prikazane u tablicama 5-39 i 5-40.



Slika 5-18. ATA 34 – prikaz vrijednosti  $F$ -mjere za  $Ac_1$

Iz slike 5-16 je vidljivo kako je maksimalna vrijednost  $F$ -mjere (od 100%) postignuta već za slučaj kada je model izgrađen slučajnim uzorkovanjem podataka i kada je udio značajki smanjen na 30% primjenom IGR filterskog postupka. Također, modeli izgrađeni slučajnim uzorkovanjem podataka kada je udio značajki smanjen na 50% i 60% temeljem KP filterskog postupka i na 40% temeljem GI postupka rezultiraju maksimalnim vrijednostima  $F$ -mjere.

Tablica 5-39. ATA 34 – prikaz vrijednosti mjere osjetljivosti za  $Ac_1$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,4286	0,5000	0,6250	0,0000	0,4286	0,5000	0,6250	0,0000	0,8571	<b>1,0000</b>	1,0000	0,6667	0,4286	0,5000	0,6250	0,0000
30	0,5000	0,7500	0,7143	0,3333	0,8571	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667	0,8571	<b>1,0000</b>	1,0000	0,6667
40	0,7143	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667	0,8333	<b>1,0000</b>	1,0000	0,6667
50	0,8333	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667	0,8333	<b>1,0000</b>	1,0000	0,6667
60	0,7143	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667
70	0,8333	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667
80	0,8333	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667	0,7143	<b>1,0000</b>	1,0000	0,6667

Tablica 5-40. ATA 34 – prikaz vrijednosti mjere specifičnosti za  $Ac_1$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,9889	1,0000	0,9875	1,0000	0,9875	1,0000	0,9875	1,0000	0,9889	0,9714	0,9875	1,0000	0,9875	1,0000	0,9875	1,0000
30	0,9889	1,0000	0,9857	1,0000	0,9875	1,0000	0,9875	1,0000	0,9889	0,9714	0,9857	1,0000	0,9889	0,9714	0,9875	1,0000
40	0,9875	0,9714	0,9857	1,0000	0,9875	0,9714	0,9857	1,0000	0,9875	0,9714	0,9857	1,0000	0,9764	1,0000	0,9857	1,0000
50	0,9764	1,0000	0,9857	1,0000	0,9875	0,9714	0,9875	1,0000	0,9875	0,9714	0,9875	1,0000	0,9875	0,9714	0,9857	1,0000
60	0,9875	1,0000	0,9857	1,0000	0,9875	0,9714	0,9875	1,0000	0,9875	0,9714	0,9875	1,0000	0,9875	0,9714	0,9875	1,0000
70	0,9875	0,9714	0,9889	1,0000	0,9875	0,9714	0,9875	1,0000	0,9875	0,9714	0,9875	1,0000	0,9875	0,9714	0,9875	1,0000
80	0,9875	0,9714	0,9889	1,0000	0,9875	0,9714	0,9875	1,0000	0,9875	0,9714	0,9875	1,0000	0,9875	0,9714	0,9875	1,0000

Iz tablice 5-39 vidljivo je kako je maksimalna vrijednost mjere *osjetljivosti* postignuta gotovo kod svih modela koji su izgrađeni slučajnim uzorkovanjem, a udio ulaznih značajki je smanjen na minimalno 40% neovisno o filterskom postupku.

Vrijednosti mjere *specifičnosti* unutar tablice 5-40 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih izgrađenih modela neovisno o udjelu ulaznih značajki i načinu uzorkovanja podataka.

#### 5.3.4.2. *Prikaz i evaluacija rezultata modela za zrakoplov Ac2*

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_2$  zabilježeno je 290 uzoraka, od čega je 271 uzoraka pripalo klasi 0, dok je preostalih 19 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 203 uzoraka, od kojih je 13 svrstano u klasu 1, a 190 u klasu 0. U skupu za verifikaciju (30%) svrstano je 87 uzoraka, od kojih je 81 pripalo klasi 0, a 6 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-41.

Tablica 5-41. *ATA 34 zrakoplov Ac2 - vrijednosti F-mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka*

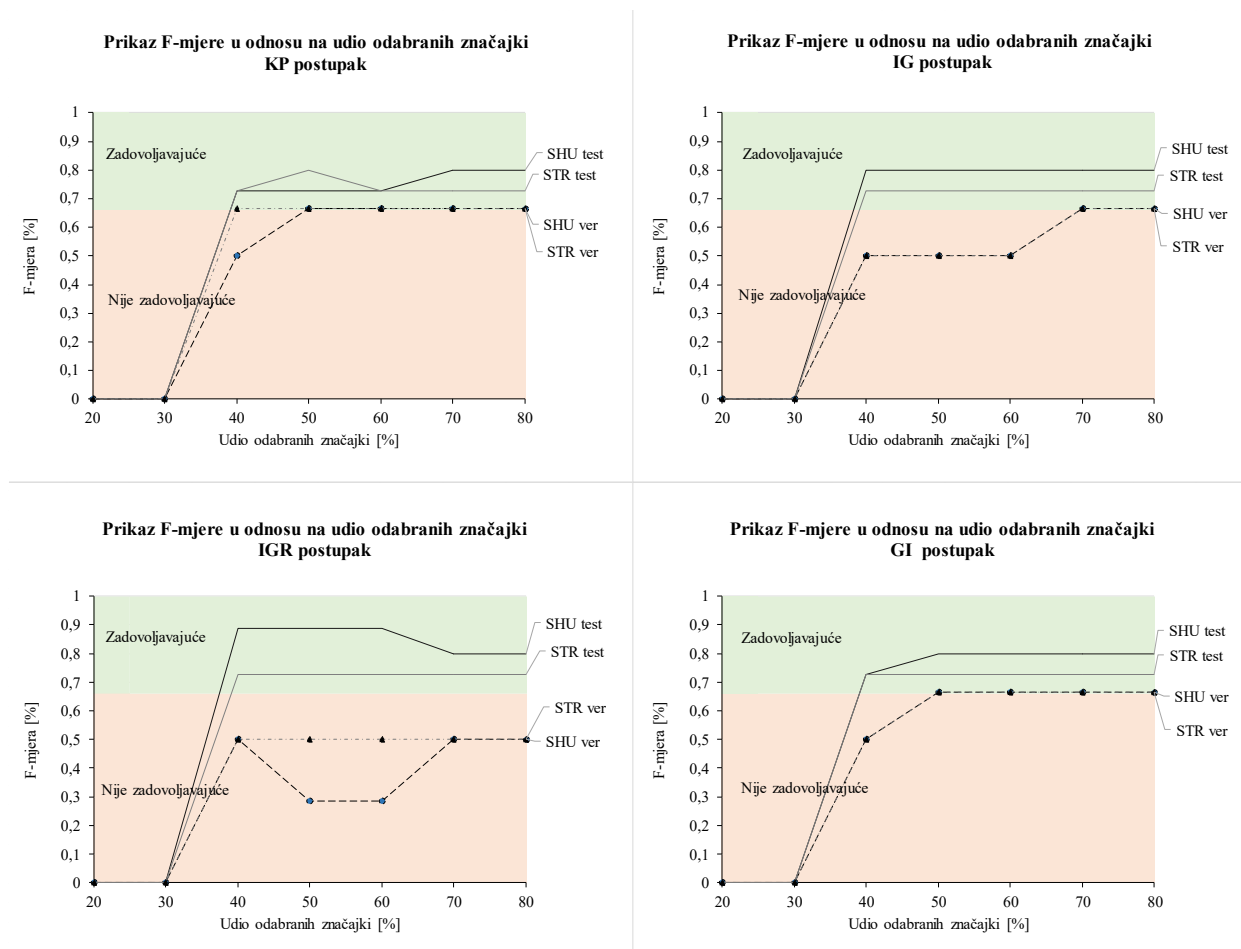
	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,8000	0,6667	0,7273	<b>0,6667</b>
<b>Osjetljivost*</b>	0,8000	0,5000	0,6667	<b>1,0000</b>
<b>Specifičnost**</b>	0,9833	1,0000	0,9857	<b>0,9630</b>

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-41 je vidljivo kako klasifikacijski model izgrađen stratificiranim uzorkovanjem podataka (STR) rezultira boljim vrijednostima statističkih mjera u odnosu na model izgrađen slučajnim uzorkovanjem podataka (SHU) te je taj model je uzet kao reprezentativan. Na slici 5-19 prikazane su vrijednosti *F-mjere* dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano rang značajki  $R_{KMIi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ . Prag za *F-mjeru* je postavljen na 0,66 (66%). Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki novoizgrađeni model su prikazane u tablicama 5-42 i 5-43.





Slika 5-19. ATA 34 – prikaz vrijednosti  $F$ -mjere za  $Ac_2$

Iz slike 5-19 vidljivo je kako smanjenjem udjela ulaznih značajki neovisno o filterskom postupku, vrijednosti  $F$ -mjere neće prijeći zadani prag. Ista vrijednost te mjere je postignuta već za slučaj kada je udio ulaznih značajki je smanjen na 40% primjenom KP filterskog postupka i model izgrađen stratificiranim uzorkovanjem podataka. Ta vrijednost ostaje nepromijenjena i za preostale udjele ulaznih značajki veće od 40% uz primjenu istog filterskog postupka i načina uzorkovanja podataka. Iste vrijednosti mjere dobivene su modelima izgrađenima neovisno o načinu uzorkovanja podataka, s udjelima ulaznih značajki smanjenim na 50%, 60%, 70% i 80% temeljem GI postupka i 70% i 80% temeljem IG postupka.

Tablica 5-42. ATA 34 – prikaz vrijednosti mjere osjetljivosti za  $Ac_2$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
30	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
40	0,6667	0,5000	0,6667	0,5000	0,8000	0,5000	0,6667	<b>1,0000</b>	0,8000	0,5000	0,6667	<b>1,0000</b>	0,8000	0,5000	0,6667	<b>1,0000</b>
50	0,6667	0,5000	0,6667	0,5000	0,8000	0,5000	0,6667	<b>1,0000</b>	0,8000	0,5000	0,6667	<b>1,0000</b>	0,8000	0,5000	0,6667	<b>1,0000</b>
60	0,6667	0,5000	0,6667	0,5000	0,8000	0,5000	0,6667	<b>1,0000</b>	0,8000	0,5000	0,6667	<b>1,0000</b>	0,8000	0,5000	0,6667	<b>1,0000</b>
70	0,8000	0,5000	0,6667	<b>1,0000</b>	0,8000	0,5000	0,6667	<b>1,0000</b>	0,8000	0,5000	0,6667	<b>1,0000</b>	0,8000	0,5000	0,6667	<b>1,0000</b>
80	0,8000	0,5000	0,6667	<b>1,0000</b>	0,8000	0,5000	0,6667	<b>1,0000</b>	0,8000	0,5000	0,6667	<b>1,0000</b>	0,8000	0,5000	0,6667	<b>1,0000</b>

Tablica 5-43. ATA 34 – prikaz vrijednosti mjere specifičnosti za  $Ac_2$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,8554	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
30	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
40	0,9833	0,9655	0,9857	1,0000	1,0000	0,9655	0,9833	0,9259	0,9833	0,9655	0,9857	0,9259	0,9714	0,9655	0,9833	0,9259
50	0,9875	1,0000	1,0000	1,0000	1,0000	0,8621	0,9833	0,9259	0,9857	0,9655	0,9857	0,9259	0,9833	1,0000	0,9857	0,9630
60	0,9833	1,0000	0,9857	0,9630	1,0000	0,8621	0,9833	0,9259	0,9833	0,9655	0,9857	0,9259	0,9833	1,0000	0,9857	0,9630
70	0,9833	1,0000	0,9857	0,9630	0,9833	0,9655	0,9857	0,9259	0,9833	1,0000	0,9857	0,9630	0,9833	1,0000	0,9857	0,9630
80	0,9833	1,0000	0,9857	0,9630	0,9833	0,9655	0,9857	0,9259	0,9833	1,0000	0,9857	0,9630	0,9833	1,0000	0,9857	0,9630

Za vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-42 vidljivo je kako smanjenjem udjela ulaznih značajki na 40%, 50%, 60%, 70% i 80%, uz primjenu IGR, IG i GI filterskih postupaka i stratificiranog uzorkovanja podataka, novoizgrađeni modeli postižu maksimalne vrijednosti mjere *osjetljivosti*. Maksimalna vrijednost mjere *osjetljivosti* se također postiže kada je udio ulaznih značajki smanjen na 70% i 80% temeljem KP postupka i model izgrađen stratificiranim uzorkovanjem podataka.

Vrijednosti mjere *specifičnosti* unutar tablice 5-43 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih izgrađenih modela neovisno o udjelu ulaznih značajki i načinu uzorkovanja podataka.

#### 5.3.4.3. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_3$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_3$  zabilježeno je 570 uzoraka, od čega je 548 uzoraka pripalo klasi 0, dok je preostalih 22 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 399 uzoraka, od kojih je 15 svrstano u klasu 1, a 384 u klasu 0. U skupu za verifikaciju (30%) svrstano je 171 uzoraka, od kojih je 164 pripalo klasi 0, a 7 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-44.

Tablica 5-44. ATA 34 zrakoplov  $Ac_3$  - vrijednosti  $F$ -mjere, *osjetljivosti* i *specifičnosti* za SHU i STR uzorkovanje podataka

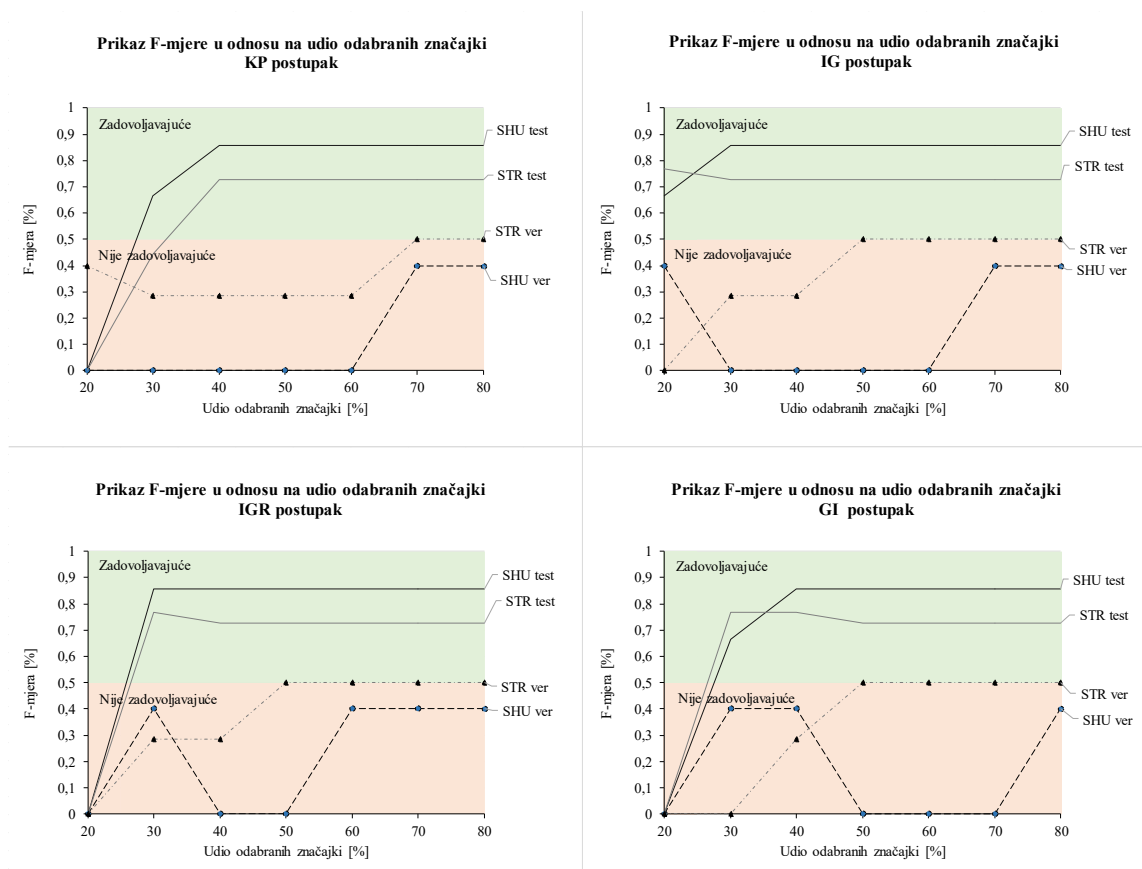
	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,7692	<b>1,0000</b>	0,9333	0,8000
<b>Osjetljivost*</b>	0,7143	<b>1,0000</b>	1,0000	0,6667
<b>Specifičnost**</b>	0,9875	<b>1,0000</b>	0,9875	1,0000

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-44 je vidljivo kako klasifikacijski model izgrađen stratificiranim uzorkovanjem podataka (STR) rezultira boljim vrijednostima  $F$ -mjere i *osjetljivosti* u odnosu na model izgrađen slučajnim uzorkovanjem podataka (SHU) te je taj model je uzet kao reprezentativan. Na slici 5-20 prikazane su vrijednosti  $F$ -mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitanoj ranga značajki  $R_{KMI}$ , gdje je  $i =$

KP, IG, IGR i GI. Prag za  $F$ -mjeru je postavljen na 0,5 (50%). Vrijednosti mjera osjetljivosti i specifičnosti za svaki novoizgrađeni model su prikazane u tablicama 5-45 i 5-46.



Slika 5-20. ATA 34 – prikaz vrijednosti  $F$ -mjere za  $Ac_3$

Iz slike 5-20 vidljivo je kako smanjenjem udjela ulaznih značajki neovisno o filterskom postupku, vrijednosti  $F$ -mjere neće prijeći zadani prag. Ista vrijednost te mjere je postignuta već za slučaj kada je udio ulaznih značajki je smanjen na 50% primjenom IG, IGR i GI filterskog postupka i model izgrađen stratificiranim uzorkovanjem podataka. Ta vrijednost ostaje nepromijenjena i za ostale udjele značajki veće od 50%, kada se primjenjuje isti filterski postupak i stratificirano uzorkovanje podataka. Također, modeli izgrađeni stratificiranim uzorkovanjem sa smanjenim udjelom značajki od 70% i 80% uz primjenu KP filterskog postupka postižu iste vrijednosti  $F$ -mjere.

Tablica 5-45. ATA 34 – prikaz vrijednosti mjere osjetljivosti za  $Ac_3$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,0000	0,0000	0,0000	0,2500	0,0000	0,0000	0,0000	0,0000	0,5000	0,2500	0,6250	0,0000	0,0000	0,0000	0,0000	0,0000
30	0,5000	0,0000	0,2857	0,2500	0,7500	0,2500	0,6250	0,2500	0,7500	0,0000	0,5714	0,2500	0,5000	0,2500	0,6250	0,0000
40	0,7500	0,0000	0,5714	0,2500	0,7500	0,0000	0,5714	0,2500	0,7500	0,0000	0,5714	0,2500	0,7500	0,2500	0,6250	0,2500
50	0,7500	0,0000	0,5714	0,2500	0,7500	0,0000	0,5714	<b>0,5000</b>	0,7500	0,0000	0,5714	<b>0,5000</b>	0,7500	0,0000	0,5714	<b>0,5000</b>
60	0,7500	0,0000	0,5714	0,2500	0,7500	0,2500	0,5714	<b>0,5000</b>	0,7500	0,0000	0,5714	<b>0,5000</b>	0,7500	0,0000	0,5714	<b>0,5000</b>
70	0,7500	0,2500	0,5714	<b>0,5000</b>	0,7500	0,2500	0,5714	<b>0,5000</b>	0,7500	0,2500	0,5714	<b>0,5000</b>	0,7500	0,0000	0,5714	<b>0,5000</b>
80	0,7500	0,2500	0,5714	<b>0,5000</b>	0,7500	0,2500	0,5714	<b>0,5000</b>	0,7500	0,2500	0,5714	<b>0,5000</b>	0,7500	0,2500	0,5714	<b>0,5000</b>

Tablica 5-46. ATA 34 – prikaz vrijednosti mjere specifičnosti za  $Ac_3$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	1,0000	1,0000	0,9818	1,0000	1,0000	1,0000	0,9364	1,0000	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	0,9364	1,0000
30	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556
40	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556
50	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556
60	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556
70	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556
80	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556	1,0000	1,0000	1,0000	0,9556

Za vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-45 vidljivo je kako smanjenjem udjela ulaznih značajki na 50%, 60%, 70% i 80%, uz primjenu IGR, IG i GI filterskih postupaka i stratificiranog uzorkovanja podataka, novoizgrađeni modeli postižu jednake vrijednosti mjere *osjetljivosti* kao i model izgrađenim s ukupnim brojem ulaznih značajki (tablica 5-44). Jednaku vrijednost mjere *osjetljivosti* postižu i modeli izgrađeni stratificiranim uzorkovanjem podataka sa smanjenim udjelom ulaznih značajki na 70% i 80% uz primjenu KP filterskog postupka.

Vrijednosti mjere *specifičnosti* unutar tablice 5-46 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih izgrađenih modela neovisno o udjelu ulaznih značajki i načinu uzorkovanja podataka.

#### 5.3.4.4. Prikaz i evaluacija rezultata modela za zrakoplov $Ac_4$

Transformacijom i balansiranjem  $D_{KMI}$  skupa podataka, u skup podataka  $D_{KMI-1}$  zrakoplova  $Ac_4$  zabilježeno je 622 uzoraka, od čega je 592 uzoraka pripalo klasi 0, dok je preostalih 30 pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 435 uzoraka, od kojih je 21 svrstano u klasu 1, a 414 u klasu 0. U skupu za verifikaciju (30%) svrstano je 187 uzoraka, od kojih je 178 pripalo klasi 0, a 9 klasi 1.

Rezultati klasifikacijskog modela izgrađenog s ukupnim brojem ulaznih značajki su prikazani u tablici 5-47.

Tablica 5-47. ATA 34 zrakoplov  $Ac_4$  - vrijednosti  $F$ -mjere, osjetljivosti i specifičnosti za SHU i STR uzorkovanje podataka

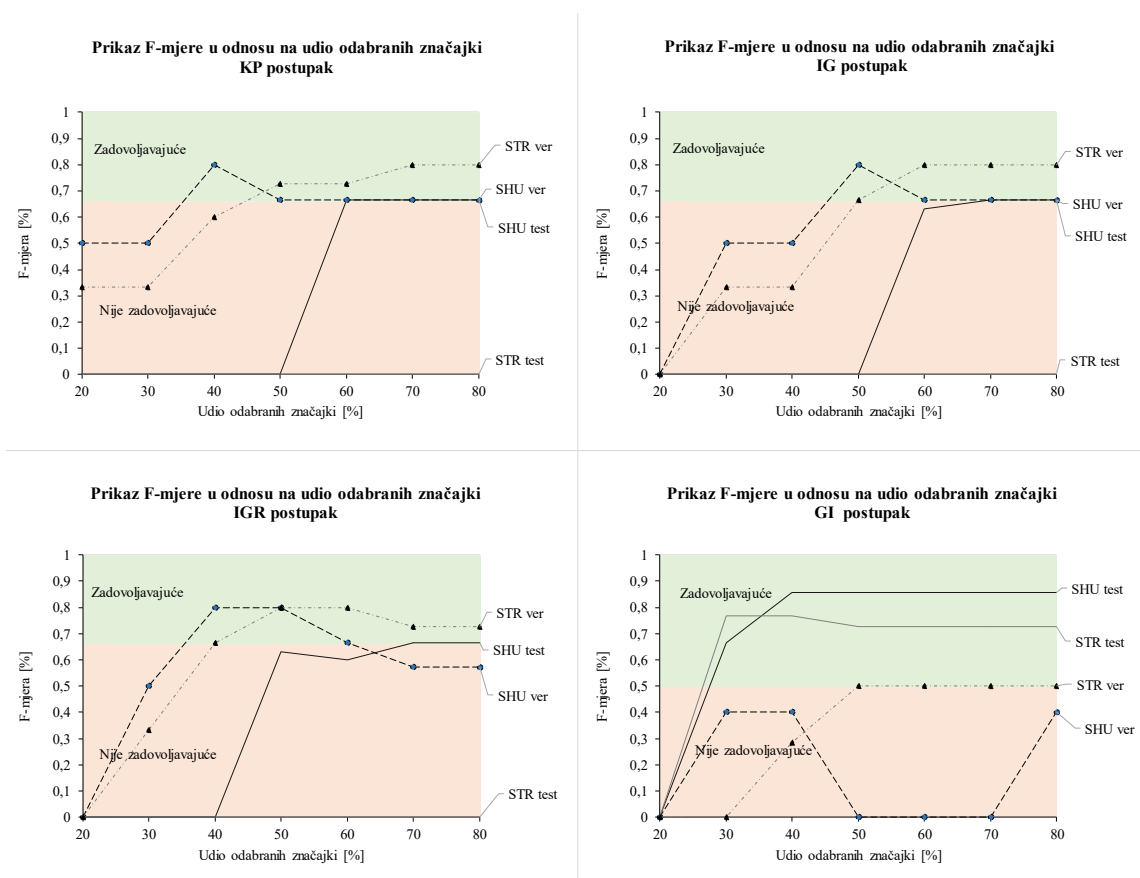
	SHU		STR	
	Test	Verifikacija	Test	Verifikacija
<b>F-mjera*</b>	0,7692	<b>1,0000</b>	0,9333	0,8000
<b>Osjetljivost*</b>	0,7143	<b>1,0000</b>	1,0000	0,6667
<b>Specifičnost**</b>	0,9875	<b>1,0000</b>	0,9875	1,0000

\* mjera za klasu 1

\*\* mjera za klasu 0

Analizirajući rezultate statističkih mjera dobivene verifikacijskim skupom podataka, iz tablice 5-47 je vidljivo kako klasifikacijski model izgrađen slučajnim uzorkovanjem podataka (SHU) ima bolje vrijednosti statističkih mjera u odnosu na model izgrađen stratificiranim uzorkovanjem podataka (STR) te je taj model uzet kao reprezentativan. Na slici 5-21 prikazane su vrijednosti  $F$ -mjere dobivene klasifikacijskim modelima kada su filtrirane ulazne značajke u udjelu od 20% do 80% temeljem učitano g ranga značajki  $R_{KMIi}$ , gdje je  $i = KP, IG, IGR$  i  $GI$ .

Prag za  $F$ -mjeru je postavljen na 0,66 (66%). Vrijednosti mjera *osjetljivosti* i *specifičnosti* za svaki od novoizgrađenih modela su prikazane u tablicama 5-48 i 5-49.



Slika 5-21. ATA 34 – prikaz vrijednosti  $F$ -mjere za  $Ac_4$

Iz slike 5-21 vidljivo je kako modeli izgrađeni slučajnim uzorkovanjem podataka postižu bolje vrijednosti  $F$ -mjere kada je udio značajki smanjen već na 40% temeljem KP i IGR filterskog postupka, no pri tome su vrijednosti  $F$ -mjere na testnim skupovima podataka loše te ti modeli nisu reprezentativni.

Iste vrijednosti  $F$ -mjere su postignute za slučaj kada su modeli izgrađeni slučajnim uzorkovanjem podataka sa smanjenim udjelom značajki na 60%, 70% i 80% temeljem KP i IG postupka i stratificiranim uzorkovanjem podataka sa smanjenim udjelom značajki od 50% do 80% temeljem GI postupka.

Tablica 5-48. ATA 34 – prikaz vrijednosti mjere osjetljivosti za  $Ac_4$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,0000	0,3333	0,0000	0,2500	0,0000	0,0000	0,0000	0,0000	0,0000	0,3333	0,0000	0,2500	0,0000	0,0000	0,0000	0,0000
30	0,0000	0,3333	0,0000	0,2500	0,0000	0,3333	0,0000	0,2500	0,0000	0,3333	0,0000	0,2500	0,0000	0,3333	0,0000	0,2500
40	0,0000	0,6667	0,0000	0,7500	0,0000	0,6667	0,0000	0,7500	0,0000	0,6667	0,0000	0,7500	0,0000	0,3333	0,0000	0,2500
50	0,0000	0,6667	0,0000	1,0000	0,6000	<b>0,6667</b>	0,0000	1,0000	0,6000	<b>0,6667</b>	0,0000	1,0000	0,0000	0,6667	0,0000	0,7500
60	0,6000	<b>0,6667</b>	0,0000	1,0000	0,6000	<b>0,6667</b>	0,0000	1,0000	0,6000	<b>0,6667</b>	0,0000	1,0000	0,6000	<b>0,6667</b>	0,0000	1,0000
70	0,6000	<b>0,6667</b>	0,0000	1,0000	0,6000	<b>0,6667</b>	0,0000	1,0000	0,6000	<b>0,6667</b>	0,0000	1,0000	0,6000	<b>0,6667</b>	0,0000	1,0000
80	0,6000	<b>0,6667</b>	0,0000	1,0000	0,6000	<b>0,6667</b>	0,0000	1,0000	0,6000	<b>0,6667</b>	0,0000	1,0000	0,6000	<b>0,6667</b>	0,0000	1,0000

Tablica 5-49. ATA 34 – prikaz vrijednosti mjere specifičnosti za  $Ac_4$

% odabranih značajki	KP postupak				IGR postupak				IG postupak				GI postupak			
	SHU		STR		SHU		STR		SHU		STR		SHU		STR	
	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver	test	ver
20	0,9545	1,0000	0,9909	0,9783	0,9636	1,0000	0,9909	1,0000	0,9727	1,0000	0,9909	0,9783	0,9818	1,0000	0,9909	1,0000
30	0,9727	1,0000	0,9909	0,9783	0,9545	1,0000	0,9909	0,9783	0,9727	1,0000	0,9909	0,9783	0,9727	1,0000	0,9909	0,9783
40	0,9545	1,0000	0,9818	0,9348	0,9545	1,0000	0,9818	0,9565	0,9545	1,0000	0,9727	0,8913	0,9727	1,0000	0,9909	0,9783
50	0,9364	0,9787	0,9727	0,9348	0,9727	1,0000	0,9627	0,9565	0,9727	1,0000	0,9627	0,9565	0,9545	1,0000	0,9818	0,9565
60	0,9800	0,9783	0,9727	0,9348	0,9636	0,9778	0,9536	0,9565	0,9636	0,9778	0,9536	0,9565	0,9727	0,9783	0,9536	0,9565
70	0,9818	0,9783	0,9727	0,9565	0,9818	0,9556	0,9727	0,9348	0,9818	0,9556	0,9727	0,9348	0,9789	0,9778	0,9727	0,9565
80	0,9809	0,9783	0,9727	0,9565	0,9818	0,9556	0,9727	0,9348	0,9818	0,9556	0,9727	0,9348	0,9789	0,9778	0,9727	0,9565



Za vrijednosti mjere *osjetljivosti* prikazanih u tablici 5-48 vidljivo je kako smanjenjem udjela ulaznih značajki na 50%, 60%, 70% i 80% primjenom IGR i IG filterskog postupka i 60%, 70% i 80% primjenom KP i GI filterskog postupka, modeli izgrađeni slučajnim uzorkovanjem podataka postižu bolje vrijednosti mjere *osjetljivosti* od modela izgrađenog s ukupnim brojem ulaznih značajki.

Vrijednosti mjere *specifičnosti* unutar tablice 5-49 ukazuju da je vrijednost uvjetne vjerojatnosti ispravno klasificiranih negativnih primjeraka zadovoljavajuća kod svih izgrađenih modela neovisno o udjelu ulaznih značajki i načinu uzorkovanja podataka.

## 5.4. Rezultati klasifikacijskih modela druge grupe

Kao što je prikazano u odjeljku 4.3, ulaz u proces za izgradnju klasifikacijskih modela druge grupe je skup podataka  $D_{KM2T}$  koji obuhvaća sjedinjene zapise o prošlim defektima (DE) i operacijskim zastoјima (OZ) te na temelju ulazne značajke *Description* predviđa zavisne značajke *Utjecaj na raspoloživost* i *Posljedica*.

### 5.4.1. ATA 22

Skup podataka  $D_{KM2T}$  za sustav ATA 22 sastoji se od ukupno 164 uzoraka, od čega su klasi 0 pripala 123 uzorka, a 41 uzorak je pripao klasi 1. Skupu za učenje modela (70%) pripalo je 115 uzoraka, od kojih je 29 svrstano u klasu 1, a 86 u klasu 0. U skup za verifikaciju (30%) je svrstano 49 uzoraka, od kojih je 37 pripalo klasi 0, a 12 klasi 1.

Rezultati klasifikacijskog modela druge grupe sustava ATA 22, za slučaj kada je zavisna značajka *Utjecaj na raspoloživost*, prikazani su u tablici 5-50.

Tablica 5-50. ATA 22 – prikaz matrice zabune kada je zavisna značajka *Utjecaj na raspoloživost*

	TESTNI PODACI			VERIFIKACIJSKI PODACI			
Točnost modela: <b>97,35%</b>	stvarno 0	stvarno 1	preciznost klase	Točnost modela: <b>97,96%</b>	stvarno 0	stvarno 1	preciznost klase
predviđeno 0	<b>86</b>	3	96,63%	predviđeno 0	<b>36</b>	0	100,00%
predviđeno 1	0	<b>26</b>	100,00%	predviđeno 1	1	<b>12</b>	92,31%
osjetljivost klase	100,00%	89,66%		osjetljivost klase	97,30%	100,00%	

Analizirajući verifikacijske podatke, iz tablice 5-50 je vidljivo kako točnost modela iznosi 97,96%, što znači da je klasifikacijska točnost modela razvijenog za sustav ATA 22 veća od točnosti modela razvijenog u radu [16]. Nadalje, vidljivo je kako model ima dobre generalizacijske sposobnosti nad novim skupom podataka za slučaj klase 1 i klase 0. Iako je preciznost klase 0 bolja od preciznosti klase 1, osjetljivost klase 1 je 100%, što znači da model točno predviđa pozitivne primjerke.

Za predviđanje točne posljedice tekstualnih zapisa koji su svrstani u klasu 1, tj. razloga neraspoloživosti, razvijen je drugi klasifikacijski model. Rezultati modela su prikazani u tablici 5-51.

Tablica 5-51. ATA 22 – prikaz matrice zabune kada je zavisna značajka Posljedica

		TESTNI PODACI				VERIFIKACIJSKI PODACI					
Točnost modela: 96,44%		Stvarni broj uzoraka			PREC	Točnost modela: 95,92%		Stvarni broj uzoraka			PREC
		I	II	IV				I	II	IV	
Predviđeno	I	86	3	1	95,56%	Predviđeno	I	36	0	0	100,00%
	II	0	23	0	100,00%		II	1	10	0	90,91%
	IV	0	0	2	100,00%		IV	0	1	1	50,00%
OSJ		100,00 %	88,46 %	66,67 %		OSJ		97,30 %	90,91 %	100,00 %	

I – klasa *None*; II – klasa *Delay*; III – klasa *Cancellation*, IV – klasa *AOG*

OSJ – osjetljivost klase; PREC – preciznost klase

Iz tablice 5-51 je vidljivo kako je točnost modela neznatno lošija od točnosti modela za slučaj kada je zavisna značajka *Utjecaj na raspoloživost*. Stoga se može zaključiti da će model dobro predvidjeti koja će biti posljedica (klasa neraspoloživosti) temeljem tekstualnog opisa nepravilnosti ili kvara na sustavu. Vidljivo je kako u skupu podataka ATA 22 sustava ni jedan zapis nije rezultirao klasom *Cancellation* (III), već samo klasama *None* (I), *Delay* (II) i *AOG* (IV). Analizirajući mjeru *preciznost* na verifikacijskom skupu podataka može se zaključiti da će model biti najmanje precizan kod predviđanja zapisa koji bi trebali biti svrstani u *AOG* klasu. Budući da se radi o malom broju uzoraka koji su pripali ovoj klasi, bolje vrijednosti mjere *preciznosti* se mogu očekivati u slučaju skupa podataka s većim brojem uzoraka svrstanih u ovu klasu.

### 5.4.2. ATA 27

Skup podataka  $D_{KM2T}$  za sustav ATA 27 sastoji se od ukupno 232 uzoraka, od čega su klasi 0 pripala 132 uzorka, a 100 uzoraka je pripalo klasi 1. Skupu za učenje modela (70%) pripala su 162 uzorka, od kojih je 70 svrstano u klasu 1, a 92 u klasu 0. U skup za verifikaciju (30%) je svrstano 70 uzoraka, od kojih je 40 pripalo klasi 0, a 30 klasi 1.

Rezultati klasifikacijskog modela druge grupe sustava ATA 27, za slučaj kada je zavisna značajka *Utjecaj na raspoloživost*, prikazani su u tablici 5-52.

Tablica 5-52. ATA 27 – prikaz matrice zabune kada je zavisna značajka *Utjecaj na raspoloživost*

	TESTNI PODACI			VERIFIKACIJSKI PODACI			
Točnost modela: <b>97,50%</b>	stvarno 0	stvarno 1	preciznost klase	Točnost modela: <b>91,43%</b>	stvarno 0	stvarno 1	preciznost klase
predviđeno 0	<b>91</b>	3	96,81%	predviđeno 0	<b>38</b>	4	90,48%
predviđeno 1	1	<b>67</b>	98,53%	predviđeno 1	2	<b>26</b>	92,86%
osjetljivost klase	98,91%	95,71%		osjetljivost klase	95,00%	86,67%	

Na temelju verifikacijskih podataka, iz tablice 5-52 je vidljivo kako točnost modela iznosi 91,43%, što znači da je klasifikacijska točnost modela razvijenog za sustav ATA 27 niža od točnosti modela razvijenog u radu [16] za približno 2%.

Vidljivo je kako model ima dobre generalizacijske sposobnosti nad novim skupom podataka, iako je dio uzoraka koji pripadaju klasi 1 klasificirano pod klasu 0.

Kao i u prethodnom slučaju za sustav ATA 22, za predviđanje točne posljedice, tj. razloga neraspoloživosti (*AOG*, *Delay*, *Cancellation* ili *None*), razvijen je drugi klasifikacijski model za sustav ATA 27. Rezultati modela su prikazani u tablici 5-53.

Tablica 5-53. ATA 27 – prikaz matrice zabune kada je zavisna značajka Posljedica

		TESTNI PODACI				VERIFIKACIJSKI PODACI					
Točnost modela: 93,82%		Stvarni broj uzoraka			PREC	Točnost modela: 94,29%		Stvarni broj uzoraka			PREC
		I	II	IV				I	II	IV	
Predviđeno	I	90	3	1	95,74%	Predviđeno	I	38	2	0	95,00%
	II	2	58	3	92,06%		II	2	24	0	92,31%
	IV	0	1	4	80,00%		IV	0	0	4	100,00%
OSJ		97,83%	93,55%	50,00%		OSJ		95,00%	92,31%	100,00%	

I – klasa *None*; II – klasa *Delay*; III – klasa *Cancellation*, IV – klasa *AOG*

OSJ – osjetljivost klase; PREC – preciznost klase

Iz tablice 5-53 je vidljivo kako je točnost modela verifikacijskog skupa podataka bolja od točnosti modela verifikacijskog skupa podataka prikazanog u tablici 5-52. Osim točnosti modela, mjere *preciznosti* i *osjetljivosti* su iznad 90% za svaku klasu, što znači da izgrađeni model dobro predviđa pozitivne i negativne uzorke. Kao i za slučaj  $D_{KM2T}$  skupa podataka sustava ATA 22, ni u skupu podataka ATA 29 sustava nema uzoraka koji su svrstani u klasu *Cancellation* (III). Analizom verifikacijske matrice zabune može se zaključiti da je dio uzoraka svrstan u klasu *Delay* (II) sličan uzorcima svrstanim u klasu *None* (I), dok su uzorci svrstani u klasu *AOG* (IV) potpuno različiti od uzoraka iz preostalih klasa.

#### 5.4.3. ATA 29

Skup podataka  $D_{KM2T}$  za sustav ATA 29 sastoji se od ukupno 82 uzoraka, od čega je klasi 0 pripalo 7 uzoraka, a 75 uzorka je pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 58 uzoraka, od kojih je 53 svrstano u klasu 1, a 5 u klasu 0. U skup za verifikaciju (30%) je svrstano 24 uzoraka, od kojih su 2 pripala klasi 0, a 22 klasi 1.

Rezultati klasifikacijskog modela druge grupe sustava ATA 29, za slučaj kada je zavisna značajka *Utjecaj na raspoloživost*, prikazani su u tablici 5-54.

Tablica 5-54. ATA 29 – prikaz matrice zabune kada je zavisna značajka Utjecaj na raspoloživost

	TESTNI PODACI			VERIFIKACIJSKI PODACI			
Točnost modela: <b>98,33%</b>	stvarno 0	stvarno 1	preciznost klase	Točnost modela: <b>100,00%</b>	stvarno 0	stvarno 1	preciznost klase
predviđeno 0	4	0	100,00%	predviđeno 0	2	0	100,00%
predviđeno 1	1	53	98,15%	predviđeno 1	0	22	100,00%
osjetljivost klase	80,00%	100,00%		osjetljivost klase	100,00%	100,00%	

Iz tablice 5-54 je vidljivo kako točnost modela iznosi 100,00% za verifikacijski skup podataka sustava ATA 29. Budući da je u  $D_{KM2T}$  skupu podataka ovog sustava većina uzoraka svrstana u klasu 1, može se zaključiti da će gotovo svaki zapis o nepravilnosti ili kvaru na ovom sustavu utjecati na raspoloživost zrakoplova. Iz tog razloga je model koji predviđa posljedice tekstualni zapisa, tj. razlog neraspoloživosti, značajniji za ovaj sustav. Rezultati modela su prikazani u tablici 5-55.

Tablica 5-55. ATA 29 – prikaz matrice zabune kada je zavisna značajka Posljedica

		TESTNI PODACI				VERIFIKACIJSKI PODACI					
Točnost modela: 91,67%		Stvarni broj uzoraka			PREC	Točnost modela: 83,33%		Stvarni broj uzoraka			PREC
		I	II	IV				I	II	IV	
Predviđeno	I	4	0	0	100,00%	Predviđeno	I	2	0	0	100,00%
	II	1	40	3	90,91%		II	0	17	4	80,95%
	IV	0	1	9	90,00%		IV	0	0	1	100,00%
OSJ		80,00%	97,56%	75,00%		OSJ		100,00%	100,00%	20,00%	

I – klasa *None*; II – klasa *Delay*; III – klasa *Cancellation*, IV – klasa *AOG*

OSJ – osjetljivost klase; PREC – preciznost klase

Iz tablice 5-55 je vidljivo kako je točnost modela verifikacijskog skupa podataka niža od točnosti modela verifikacijskog skupa podataka prikazanog u tablici 5-54. Razlog tome je vidljiv iz matrice zabune kod klasifikacije uzoraka koji pripadaju *AOG* klasi (IV). Može se pretpostaviti kako je dio uzoraka u verifikacijskom skupu podataka, koji zapravo pripada klasi *AOG*, sličan uzorcima koji pripadaju klasi *Delay* (II). Stoga je i osjetljivost klase *AOG* niža od osjetljivosti preostalih klasa.

#### 5.4.4. ATA 32

Skup podataka  $D_{KM2T}$  za sustav ATA 32 sastoji se od ukupno 238 uzoraka, od čega je klasi 0 pripao 101 uzorak, a 137 uzoraka je pripalo klasi 1. Skupu za učenje modela (70%) pripalo je 167 uzoraka, od kojih je 96 svrstano u klasu 1, a 71 u klasu 0. U skup za verifikaciju (30%) je svrstan 71 uzorak, od kojih je 30 pripalo klasi 0, a 41 klasi 1.

Rezultati klasifikacijskog modela druge grupe sustava ATA 32, za slučaj kada je zavisna značajka *Utjecaj na raspoloživost*, prikazani su u tablici 5-56.

Tablica 5-56. ATA 32 – prikaz matrice zabune kada je zavisna značajka *Utjecaj na raspoloživost*

	TESTNI PODACI			VERIFIKACIJSKI PODACI			
Točnost modela: <b>96,40%</b>	stvarno 0	stvarno 1	preciznost klase	Točnost modela: <b>97,18%</b>	stvarno 0	stvarno 1	preciznost klase
predviđeno 0	<b>67</b>	2	97,10%	predviđeno 0	<b>29</b>	1	96,67%
predviđeno 1	4	<b>94</b>	95,92%	predviđeno 1	1	<b>40</b>	97,56%
osjetljivost klase	94,37%	97,92%		osjetljivost klase	96,67%	97,56%	

Temeljem verifikacijskog skupa podataka, iz tablice 5-56 je vidljivo kako točnost modela, za slučaj kada se predviđa kategorička značajka *Utjecaj na raspoloživost*, iznosi 97,18%. To znači da je klasifikacijska točnost modela razvijenog za sustav ATA 32 veća od točnosti modela razvijenog u radu [16]. Kao i u skupu podataka za sustav ATA 29, u  $D_{KM2T}$  skupu podataka za sustav ATA 32 je većina uzoraka svrstana u klasu 1 te je klasifikacijski model koji predviđa posljedicu nerasploživosti od veće važnosti za ovaj sustav. Rezultati modela, za slučaj gdje je zavisna značajka *Posljedica*, su prikazani u tablici 5-57.

Tablica 5-57. ATA 32 – prikaz matrice zabune kada je zavisna značajka Posljedica

		TESTNI PODACI					VERIFIKACIJSKI PODACI					
Točnost modela: <b>82,02%</b>		Stvarni broj uzoraka				PREC	Točnost modela: <b>73,24%</b>	Stvarni broj uzoraka				PREC
		I	II	III	IV			I	II	III	IV	
Predviđeno	I	<b>67</b>	2	0	1	95,71 %	Predviđeno	I	<b>29</b>	0	0	100,00 %
	II	3	<b>57</b>	2	9	80,28 %		II	1	<b>21</b>	0	70,00 %
	III	0	3	<b>0</b>	2	0,00 %		III	0	2	<b>0</b>	0,00 %
	IV	1	6	1	<b>13</b>	61,90 %		IV	0	6	1	22,22 %
OSJ		94,37 %	83,82 %	0,00 %	52,00 %		OSJ		96,67 %	72,41 %	0,00 %	18,18 %

I – klasa *None*; II – klasa *Delay*; III – klasa *Cancellation*, IV – klasa *AOG*

OSJ – osjetljivost klase; PREC – preciznost klase

Iz tablice 5-56 je vidljivo kako je točnost modela verifikacijskog skupa podataka niža od točnosti modela verifikacijskog skupa podataka prikazanog u tablici 5-55. Za razliku od prethodnih  $D_{KM2T}$  skupova podataka, u  $D_{KM2T}$  skupu podataka ATA 32 sustava nalaze se uzorci koji su svrstani u klasu *Cancellation* (III). Analizom matrice zabune dobivene testnim skupom podataka je vidljivo kako izgrađeni model loše predviđa uzorke svrstane u klasu *Cancellation*, što je potvrđeno i na verifikacijskom skupu podataka. Izgrađeni model također loše klasificira uzorke koji pripadaju *AOG* klasi (IV) jer je većina uzoraka slična uzorcima iz klase *Delay* (II) i *Cancellation* (III).

#### 5.4.5. ATA 34

Skup podataka  $D_{KM2T}$  za sustav ATA 34 sastoji se od ukupno 89 uzoraka, od čega su klasi 0 pripala 66 uzorka, a 23 uzorka su pripala klasi 1. Skupu za učenje modela (70%) pripalo je 62 uzorka, od kojih je 16 svrstano u klasu 1, a 46 u klasu 0. U skup za verifikaciju (30%) je svrstano 27 uzoraka, od kojih je 20 pripalo klasi 0, a 7 klasi 1.

Rezultati klasifikacijskog modela druge grupe sustava ATA 34, za slučaj kada je zavisna značajka *Utjecaj na raspoloživost*, prikazani su u tablici 5-58.

Tablica 5-58. ATA 34 – prikaz matrice zabune kada je zavisna značajka Utjecaj na raspoloživost

	TESTNI PODACI			VERIFIKACIJSKI PODACI			
Točnost modela: <b>93,24%</b>	stvarno 0	stvarno 1	preciznost klase	Točnost modela: <b>92,59%</b>	stvarno 0	stvarno 1	preciznost klase
predviđeno 0	<b>43</b>	3	93,48%	predviđeno 0	<b>20</b>	2	90,91%
predviđeno 1	1	<b>13</b>	92,86%	predviđeno 1	0	<b>5</b>	100,00%
osjetljivost klase	97,73%	81,25%		osjetljivost klase	100,00%	71,43%	

Analizirajući verifikacijske podatke, iz tablice 5-58 je vidljivo kako točnost modela iznosi 92,59%, što znači da je klasifikacijska točnost modela razvijenog za sustav ATA 34 gotovo jednaka kao i točnosti modela razvijenog u radu [16]. Iz verifikacijskog skupa podataka je također vidljivo kako je izgrađeni model preciziran kod predviđanja klase 1, no osjetljivost te klase je lošiji što znači da će točnost pozitivnih primjeraka biti manja u odnosu na točnost negativnih primjeraka. Rezultati modela, za slučaj gdje je zavisna značajka *Posljedica*, su prikazani u tablici 5-59.

Tablica 5-59. ATA 34 – prikaz matrice zabune kada je zavisna značajka Posljedica

		TESTNI PODACI				VERIFIKACIJSKI PODACI					
Točnost modela: 89,67%		Stvarni broj uzoraka			PREC	Točnost modela: 85,19%		Stvarni broj uzoraka			PREC
		I	II	IV				I	II	IV	
Predviđeno	I	43	2	0	95,56%	Predviđeno	I	20	0	2	90,91%
	II	0	6	3	66,67%		II	0	1	0	100,00%
	IV	1	0	5	83,33%		IV	0	2	2	50,00%
OSJ		97,73%	75,00%	62,50%		OSJ		100,00%	33,33%	50,00%	

I – klasa *None*; II – klasa *Delay*; III – klasa *Cancellation*, IV – klasa *AOG*

OSJ – osjetljivost klase; PREC – preciznost klase

Iz tablice 5-59 je vidljivo kako je točnost modela verifikacijskog skupa podataka niža od točnosti modela verifikacijskog skupa podataka prikazanog u tablici 5-58. Kao i za slučaj  $D_{KM2T}$  skupova podataka sustava ATA 22, ATA 27 i ATA 29, ni u skupu podataka ATA 34 sustava, nije bilo uzoraka koji su svrstani u klasu *Cancellation* (III). Analizom verifikacijske matrice zabune može se zaključiti da je dio uzoraka svrstan u klasu *Delay* (II) sličan uzorcima svrstanim



u klasu *AOG* (IV). Slični zaključci se mogu povući i za uzorke u verifikacijskom skupu koji pripadaju *AOG* klasi. Dio tih uzoraka je svrstan u klasu *None* (I) što znači da su značajke (pojmovi) unutar tih uzoraka vrlo slični značajkama (pojmovima) koji neće utjecati na raspoloživost zrakoplova.

## 5.5. Zaključak poglavlja

Iz odjeljka 5.2.2 je vidljivo kako su asocijativna pravila, prema zadanim pragovima parametara minimalne podrške od 0,1 i minimalne pouzdanosti od 0,5, pronađena jedino za skup podataka sustava ATA 29, što dodatno potvrđuje uvodne napomene o nezavisnosti ulaznih značajki u odnosu na pojavu zavisne značajke.

Iz odjeljka 5.3 je vidljivo kako je najstroža mjera za evaluaciju klasifikacijskih modela prve grupe *F*-mjera jer ona predstavlja kombinaciju mjere *osjetljivosti* (*odziva*) i *preciznosti* slabe (pozitivne) klase. Iz tog razloga, zaključci prikazani za klasifikacijske modele prve grupe će biti sažeti na temelju dobivenih vrijednosti te mjere.

Temeljem rezultata klasifikacijskih modela prve grupe, u kojima je ispitivan utjecaj različitih načina uzorkovanja podataka u kombinaciji s filterskim postupcima za smanjenje udjela ulaznih značajki, svi zaključci se mogu sažeti u sljedećim točkama:

1. Bolje ili jednake vrijednosti *F*-mjere za ATA 22 sustav postignute su sljedećim modelima:
  - a.  $\{(ATA\ 22)_{Ac1}; STR; 40\%; IGR\}$ ; za zrakoplov  $Ac_1$  kada je model izgrađen stratificiranim uzorkovanjem podataka sa smanjenim udjelom značajki na 40% temeljem IGR filterskog postupka,
  - b.  $\{(ATA\ 22)_{Ac2}; SHU; 20\%; GI\}$ ; za zrakoplov  $Ac_2$  kada je model izgrađen slučajnim uzorkovanjem podataka sa smanjenim udjelom značajki na 20% temeljem GI filterskog postupka,
  - c.  $\{(ATA\ 22)_{Ac3}; SHU; 30\%; KP\}$ ; za zrakoplov  $Ac_3$  kada je model izgrađen slučajnim uzorkovanjem podataka sa smanjenim udjelom značajki na 30% temeljem KP filterskog postupka,
  - d.  $\{(ATA\ 22)_{Ac4}; STR; 80\%; KP\}$ ; za zrakoplov  $Ac_4$  kada je model izgrađen stratificiranim uzorkovanjem podataka sa smanjenim udjelom značajki na 80% temeljem KP filterskog postupka.

2. Bolje ili jednake vrijednosti  $F$ -mjere za ATA 27 sustav postignute su sljedećim modelima:
  - a.  $\{(ATA\ 27)_{Ac1}; SHU; 50\%; KP, IG, IGR, GI\}$ ; za zrakoplov  $Ac_1$  kada je model izgrađen slučajnim uzorkovanjem podataka sa smanjenim udjelom značajki na 50% neovisno o filterskom postupku,
  - b.  $\{(ATA\ 27)_{Ac2}; SHU; 60\%; KP, IGR\}$ ; za zrakoplov  $Ac_2$  kada je model izgrađen slučajnim uzorkovanjem podataka sa smanjenim udjelom značajki na 60% temeljem KP i IGR filterskog postupka,
  - c.  $\{(ATA\ 27)_{Ac3}; STR; 70\%; IG\}$ ; za zrakoplov  $Ac_3$  kada je model izgrađen stratificiranim uzorkovanjem podataka sa smanjenim udjelom značajki na 70% temeljem IG filterskog postupka,
  - d.  $\{(ATA\ 27)_{Ac4}; STR; 20\%; KP, IG, GI\}$ ; za zrakoplov  $Ac_4$  kada je model izgrađen slučajnim uzorkovanjem podataka sa smanjenim udjelom značajki na 20% temeljem KP, IG i GI filterskog postupka.
  
3. Bolje ili jednake vrijednosti  $F$ -mjere za ATA 32 sustav postignute su sljedećim modelima:
  - a.  $\{(ATA\ 32)_{Ac1}; STR; 70\%; IG, IGR, GI\}$ ; za zrakoplov  $Ac_1$  kada je model izgrađen stratificiranim uzorkovanjem podataka sa smanjenim udjelom značajki na 70% temeljem IG, IGR i GI filterskog postupka,
  - b.  $\{(ATA\ 32)_{Ac2}; SHU; 60\%; IG, IGR\}$ ; za zrakoplov  $Ac_2$  kada je model izgrađen slučajnim uzorkovanjem podataka sa smanjenim udjelom značajki na 60% temeljem IG i IGR filterskog postupka,
  - c.  $\{(ATA\ 32)_{Ac3}; STR; 40\%; KP\}$ ; za zrakoplov  $Ac_3$  kada je model izgrađen stratificiranim uzorkovanjem podataka sa smanjenim udjelom značajki na 40% temeljem KP filterskog postupka,
  - d.  $\{(ATA\ 32)_{Ac4}; STR; 30\%; KP\}$ ; za zrakoplov  $Ac_4$  kada je model izgrađen stratificiranim uzorkovanjem podataka sa smanjenim udjelom značajki na 30% temeljem KP filterskog postupka.
  
4. Bolje ili jednake vrijednosti  $F$ -mjere za ATA 34 sustav postignute su sljedećim modelima:
  - a.  $\{(ATA\ 34)_{Ac1}; SHU; 30\%; IGR\}$ ; za zrakoplov  $Ac_1$  kada je model izgrađen slučajnim uzorkovanjem podataka sa smanjenim udjelom značajki na 30% temeljem IGR filterskog postupka,

- b.  $\{(ATA\ 34)_{Ac2}; STR; 40\%; KP\}$ ; za zrakoplov  $Ac_2$  kada je model izgrađen stratificiranim uzorkovanjem podataka sa smanjenim udjelom značajki na 40% temeljem KP filterskog postupka,
- c.  $\{(ATA\ 34)_{Ac3}; SHU; 50\%; IG, IGR, GI\}$ ; za zrakoplov  $Ac_3$  kada je model izgrađen slučajnim uzorkovanjem podataka sa smanjenim udjelom značajki na 50% temeljem IG, IGR i GI filterskog postupka,
- d.  $\{(ATA\ 34)_{Ac4}; SHU; 40\%; KP, IGR\}$ ; za zrakoplov  $Ac_4$  kada je model izgrađen slučajnim uzorkovanjem podataka sa smanjenim udjelom značajki na 40% temeljem KP i IGR filterskog postupka.

Iz prethodno prikazanih rezultata *F-mjere* može se zaključiti kako će rezultati klasifikacije biti bolji kada se primjenjuju filterski postupci za selekciju značajki. To podrazumijeva da će novoizgrađeni modeli imati bolje generalizacijske sposobnosti za klasifikaciju novih uzoraka, tj. za prepoznavanje uzoraka koji će prethoditi zapisu u TLB.

Nadalje, uspoređujući rezultate mjere *točnosti* klasifikacijskih modela prikazanih u odjeljku 5.4 s rezultatima mjere *točnosti* klasifikacijskog modela prikazanog u radu [16], uočeno je kako modeli izgrađeni u ovoj disertaciji u većini slučajeva rezultiraju boljim vrijednostima te mjere. Iz navedenog se može zaključiti kako će izgrađeni modeli dobro klasificirati tekstualne zapise s ciljem prepoznavanja da li će ti zapisi utjecati na raspoloživost zrakoplova i, ako da, koja će biti točno posljedica neraspoloživosti.

## **6. ZAKLJUČAK**

U ovome poglavlju dan je osvrt na početna istraživanja, postavljenje istraživačke ciljeve i hipotezu istraživanja. Prikazani su relevantni zaključci proizašli iz istraživanja i eksperimentalnih rezultata prikazanih u prethodnim poglavljima. Navedena su ograničenja istraživanja te su dane smjernice za nastavak istraživanja.

### **6.1. Osvrt na početna istraživanja**

Inicijalno istraživanje za razvoj klasifikacijskih modela prve grupe je bilo bazirano na pretpostavci da je moguće predvidjeti pojavu pilotskog zapisa u TLB na razini zrakoplova isključivo na temelju pojave događaja na razini sustava. Pri tome, pod događajem se smatralo da nije važno koja poruka upozorenja se pojavila, već samo da je zabilježeno upozorenje u određenoj fazi leta na određeni datum (let) i po određenom sustavu (ATA-i).

Slijedeći tu pretpostavku, podaci su transformirani na način da je ulazni skup značajki bio kombinacija svih ATA i faza leta koje su poprimile vrijednosti o broju svih upozorenja registriranih na tom letu. Zavisna značajka je i dalje predstavljala kategoričku vrijednost, tj. polje u kojemu se nalazila informacija da li se dogodio zapis u TLB na taj datum i po toj ATA-i. Iako su određeni sustavi u zrakoplovu međusobno povezani, kod transformacije podataka na ovaj način, povezani su i podaci iz sustava koji nisu fizički povezani. Izgradnjom klasifikacijskih modela prve grupe, uočeno je kako na ovaj način nije moguće predvidjeti pojavu zapisa u TLB jer ulazne značajke ne opisuju dovoljno detaljno zavisnu značajku.

Iz navedenog razloga je promijenjen smjer istraživanja te su klasifikacijski modeli prve grupe izgrađeni na razini kritičnih sustava, a ulazne značajke su predstavljale kombinaciju svih poruka upozorenja generiranih na razini sustava po određenim fazama leta.

### **6.2. Osvrt na znanstvene doprinose i hipotezu istraživanja**

Motivacija za istraživanje u ovoj disertaciji je proizašla iz složenosti i rastuće količine podataka u procesu održavanja zrakoplova, koji mogu poslužiti kao potpora pri donošenju odluka o planiranju ljudskih i materijalnih resursa i time utjecati na raspoloživost zrakoplova. Stoga su u uvodnom dijelu disertacije (v. odjeljak 1.3) definirani očekivani znanstveni doprinosi, koji su ostvareni ovim istraživanjem.

Doprinos #1; *formulacija postupka za strukturiranje podataka prikupljenih iz sustava za nadzor tehničke ispravnosti zrakoplova te postupka integracije tih podataka s drugim izvorima podataka*, je prikazan u odjeljku 3.3. Za ostvarivanje ovog doprinosa razvijena su dva algoritma (prilog A i B) kojima je omogućeno strukturiranje poruka upozorenja prikupljenih iz AHMS-a i sastavljanje skupova podataka za izgradnju klasifikacijskih modela.

Doprinos #2; *identificiranje i vrednovanje relevantnih značajki po svim fazama leta na razini kritičnih sustava zrakoplova*, je prikazan u odjeljku 5.2.1. Za ostvarivanje ovog doprinosa razvijen je proces za otkrivanje relevantnih značajki primjenom RapidMiner alata za dubinsku analizu podataka (odjeljak 4.1 i prilog C). Značajke su identificirane i vrednovane korištenjem filterskih postupaka, čime je dobiven rang ulaznih značajki za svaki analizirani sustav (prilog D).

Doprinos #3; *izrada novih klasifikacijskih modela za pouzdanije predviđanje ponašanja kritičnih sustava zrakoplova*, je prikazan u odjeljcima 5.3 i 5.4. Za ostvarivanje ovog doprinosa razvijeni su procesi za izgradnju klasifikacijskih modela prve i druge grupe primjenom RapidMiner alata za dubinsku analizu podataka (odjeljak 4.2 i 4.3). Ovaj doprinos je ostvaren nakon što su strukturirani podaci iz AHMS-a i sastavljeni skupovi podataka (ostvarivanjem doprinosa #1) te stvoren rang relevantnih značajki (ostvarivanjem doprinosa #2).

Temeljem realiziranih istraživačkih ciljeva i rezultata prezentiranih u poglavlju 5, postavljena hipoteza istraživanja; *otkrivanjem značajki koje opisuju ponašanje kritičnih sustava zrakoplova u različitim fazama leta zrakoplova, na temelju tehnika dubinske analize podataka, moguće je predvidjeti raspoloživost zrakoplova*, je **potvrđena**.

Hipoteza istraživanja je potvrđena prvenstveno na temelju rezultata klasifikacijskih modela prve grupe. Vrijednosti evaluacijskih mjera (*F-mjere*, *osjetljivosti* i *specifičnosti*) tih modela ukazuju da je moguće predvidjeti uzorke poruka upozorenja po fazama leta koji prethode pilotskom upisu u TLB. Budući da pilotski zapis može utjecati na raspoloživost zrakoplova, otkrivanjem tih uzoraka i poduzimanjem prijevremenih akcija je moguće osigurati da će zrakoplov biti dostupan za obavljanje svoje funkcije prema inicijalnom planu leta. Promatrajući vrijednost *F-mjere* verifikacijskih skupova podataka, za slučaj kada su modeli izgrađeni s ukupnim brojem ulaznih značajki, ustanovljeno je da su najbolji rezultati postignuti na  $D_{KMI-1}$  skupovima podataka svih  $Ac_i$  zrakoplova kod sustava ATA 27 (iznad 0,66, osim za slučaj  $Ac_3$ ) i ATA 34 (1,00, osim za slučaj  $Ac_2$ ). Vrijednosti mjere *osjetljivosti* su kod gotovo svih  $D_{KMI-1}$  verifikacijskih skupova podataka bolje od vrijednosti *F-mjere* (osim za slučaj  $Ac_4$  sustava ATA

22,  $Ac_1$  sustava ATA 27,  $Ac_4$  sustava ATA 32), iz čega se može zaključiti da će model točno predviđati pozitivne primjerke, no neće uvijek biti precizan prilikom predviđanja. Drugim riječima, modeli točno predviđaju primjerke pozitivne klase (slabe klase), no dio primjeraka iz negativne klase (jake klase) je klasificirano unutar pozitivne klase (to su lažno pozitivni primjerci). Za domenski slučaj to ne predstavlja problem jer je bitnija točnost predviđanja primjeraka slabe klase (uzoraka koji će prethoditi upisu u TLB) od preciznosti predviđanja. Zbog većeg broja negativnih primjeraka u skupovima podataka, iz rezultata prikazanih u odjeljku 5.3, vidljivo je kako modeli rezultiraju dobrim vrijednostima mjere *specifičnosti* (iznad 0,70, osim za slučaj  $Ac_2$  sustava ATA 32). Iz toga se može zaključiti da je za dane skupove podataka mali broj primjeraka iz negativne klase bio sličan primjercima iz pozitivne klase.

Iako evaluacijske mjere klasifikacijskih modela prve grupe ukazuju na to da razvijeni modeli imaju dobre generalizacijske sposobnosti za razvrstavanje novih uzoraka, tim modelima nije moguće predvidjeti apsolutno svaki zapis u TLB. Stoga su dodatno izgrađeni klasifikacijski modeli druge grupe. Evaluacijska mjera *točnosti* je odabrana kao glavna mjera za usporedbu izgrađenih modela s postojećim (sličnim) modelom iz rada [16]. Vrijednosti mjere *točnosti* iz modela razvijenih u ovom radu su, za slučaj predviđanja ciljne značajke *Utjecaj na raspoloživost*, kod gotovo svih  $DKM2T$  skupova podataka približno jednake ili veće od 93%, tj. veće od *točnosti* dobivene modelom iz rada [16]. Za slučaj predviđanja ciljne značajke *Posljedica*, klasifikacijski modeli za sustave ATA 22 i ATA 27 su rezultirali boljim vrijednostima mjere *točnosti* (iznad 94%). Modeli razvijeni za ostale sustave su rezultirali lošijim vrijednostima mjere *točnosti* (ATA 29  $\approx$  83%, ATA 32  $\approx$  73% i ATA 34  $\approx$  85%), iz čega se može zaključiti da unutar tih skupova podataka postoji sličnost između zapisa koji pripadaju klasama *Delay*, *Cancellation* i *AOG*. Iako su rezultati mjere *točnosti* većinom niži kod modela kada se predviđa *Posljedica*, i dalje predstavljaju zadovoljavajuće vrijednosti (iznad 70%). Time je dodatno potvrđena hipoteza istraživanja i dokazana pretpostavka da je temeljem tehnika dubinske analize podataka moguće predvidjeti raspoloživost zrakoplova.

### 6.3. Teorijske i praktične implikacije

Teorijske implikacije u ovoj disertaciji proizašle su iz pregleda objavljenih publikacija u području održavanja zrakoplova vezanih uz tehnike dubinske analize podataka (poglavlje 1 i 2). Ustanovljeno je da tehnike dubinske analize podataka ne rezultiraju nužno izravnim rješenjima i odgovorima kod specifičnih domenskih problema (npr. točno vrijeme kada će se dogoditi

zastoj), zbog čega se većinom koriste kao sustavi podrške pri odlučivanju i mogu se kombinirati s drugim metodama kao što su sustavi poslovne inteligencije.

Osim navedenog, analizom literature su otkriveni i glavni nedostaci u domenskom području; nedostatan korištenje podataka prikupljenih iz AHMS-a, nedostatak primjene filterskih tehnika za selekciju relevantnih značajki te odabir klasifikacijskih algoritama i statističkih mjera bez dubljeg razumijevanja podataka.

Temeljem teorijskih uvida postavljeni su istraživački ciljevi sa svrhom uklanjanja tih nedostataka te su razvijeni novi modeli. Razvijeni modeli, bazirani na tehnikama dubinske analize podataka, su rezultirali i stručnim implikacijama.

Analizom podataka i otkrivanjem znanja na način prikazan u ovoj disertaciji napravljen je odmak od trenutnog načina analize podataka u praktičnoj primjeni. Većina trenutnih analiza, koje zračni prijevoznik provodi, vezana je uz izvještaje o pouzdanosti zrakoplovne flote (v. odjeljak 3.1.3) i bazirana je na manjem skupu podataka, tj. temeljem podataka prikupljenih u razdoblju od jednog mjeseca.

Nadalje, zbog nestrukturiranog zapisa podataka prikupljenih iz AHMS-a, te podatke nije moguće dovesti u izravnu korelaciju s drugim podacima (zapisima o nepravilnostima u radu sustava/kvarovima). Stoga je razvojem algoritma za strukturiranje tih zapisa (prilog A) omogućena analiza novih podataka.

Na temelju prikazanog istraživanja u ovoj disertaciji, razvijeni klasifikacijski modeli se mogu primijeniti i van domene zrakoplovne industrije, poput nuklearne industrije, gdje je proces nadziranja kritičnih sustava i bilježenja svih informacija o zahvatima održavanja vrlo sličan.

#### **6.4. Ograničenja istraživanja**

Ograničenja istraživanja u ovoj disertaciji su povezana s algoritmom za prepoznavanje pilotskih zapisa u TLB-u nastalih temeljem poruka upozorenja i razvijenim klasifikacijskim modelima prve grupe.

Na rješenje razvijenog algoritma prikazanog u prilogu B, koji omogućava prepoznavanje uzoraka iz skupa zapisa o prošlim nepravilnostima u radu/kvarovima sustava ( $D_{DE}$ ) nastalih na temelju uzoraka iz skupa zapisa o porukama upozorenja ( $D_{WN}^T$ ), uvelike utječe ljudski faktor. Budući da su analizirani zapisi iz skupa  $D_{DE}$  uneseni naknadno u relacijsku bazu podataka

temeljem pilotskih upisa u TLB, ti zapisi su podložni ljudskim pogreškama. Drugim riječima, kod unošenja podataka u bazu može se dogoditi pogrešan unos određenog polja, kao što su datum ili ATA zabilježenog kvara ili nepravilnost u radu sustava. Ukoliko je datum ili ATA zapisa iz pilotskog izvora pogrešno unesena u bazu, algoritam nije prepoznao da je taj zapis nastalo na temelju poruke upozorenja i zavisna značajka *Zapis* u skupu podataka  $D_{WN}^T$  nije poprimila vrijednost 1 (v. odjeljak 3.2.1).

Drugo ograničenje u istraživanju je vezano uz klasifikacijski model prve grupe. Razvijeni klasifikacijski modeli prve grupe služe za predviđanje uzoraka poruka upozorenja koje neće rezultirati automatskim upisom u TLB, tj. na manje kritične poruke koje se ponavljaju više puta prije upisa u TLB. Ovaj model nije primjenjiv na sustave gdje gotovo svaka poruka upozorenja rezultira upisom u TLB, kao što je to slučaj za hidraulički sustav (ATA 29).

## 6.5. Smjernice za daljnja istraživanja

Na temelju teorijskih i praktičnih implikacija te ograničenja istraživanja, smjernice za buduća istraživanja su vezane uz sastavljanje novih skupova podataka te korištenje hibridnih tehnika dubinske analize podataka.

U istraživanju prikazanom u ovoj disertaciji izravno su povezani skupovi podataka WN i DE (za izgradnju klasifikacijskih modela prve grupe) te skupovi DE i OZ (za izgradnju klasifikacijskih modela druge grupe) (odjeljak 3.3). Vidljivo je kako skupovi WN i OZ nisu izravno povezani i dodatno analizirani. Ti skupovi su indirektno povezani preko skupa DE jer je, za potrebe izrade klasifikacijskih modela druge grupe, skup OZ udružen sa zapisima iz skupa DE koji su nastali isključivo temeljem WN zapisa. Buduće istraživanje se predlaže upravo u tom smjeru, a to je udruživanje i analiziranje WN i OZ skupova podataka te provođenje analize na razini kritičnih komponenata u sustavu.

Osim sastavljanja skupa podataka koji bi udruživao WN i OZ skupove, za buduća istraživanja bi bilo zanimljivo analizirati skupove podataka WN i DE za sve sustave koji su međusobno povezani. Na primjer, poznato je kako hidraulički sustav pogoni niz komponenata unutar drugih, međusobno povezanih, sustava (v. odjeljak 5.2.1.3). To podrazumijeva da će kvar/nepravilnost u radu tog sustava utjecati i na funkcije drugih sustava ili komponenata unutar tih sustava. Stoga bi se udruživanjem skupova podataka WN i DE iz više različitih sustava, za izgradnju klasifikacijskih modela prve grupe, omogućio holistički pristup kod predviđanja pilotskih upisa u TLB.



Također, jedan od budućih smjerova istraživanja je vezan uz analizu dijela zapisa poruka upozorenja gdje se bilježe informacije s kojeg računala je poruka upozorenja prikupljena (v. odjeljak 3.1.1, sliku 3-1). Na ovaj način bi se dodatno pročistio skup podataka poruka upozorenja jer je na PFR-u zabilježeno niz poruka upozorenja bez informacija koje računalo je registriralo tu poruku. To znači da se takve poruke upozorenja, za koje nije zabilježena lokacija prikupljanja, mogu smatrati „lažnima“ (eng. *spurious*). U ovoj disertaciji je otkrivanje „lažnih“ poruka upozorenja prepušteno filterskim postupcima (kod rangiranja relevantnih značajki) i klasifikacijskom algoritmu (kod prepoznavanja uzoraka koji prethode upisu u TLB).

Nadalje, temeljem znanja usvojenih od strane stručnjaka za održavanje, otkriveno je da u nekim slučajevima iste poruke upozorenja mogu biti povezane s nepravilnostima u radu/kvarovima različitih komponenata unutar sustava. Samim time, pojava poruka upozorenja koje mogu biti rezultat kvara različitih komponenta otežava postupak identifikacije grešaka. Ovaj postupak se može unaprijediti korištenjem kombinacije dubinske analize teksta i nenadziranog učenja (klaster analize). Unutar skupa podataka koji sadržava zapise o kvarovima/nepravilnostima u radu sustava, dubinskom analizom teksta je moguće otkriti glavne uzroke kvarova na sustavu, tj. kritične komponente. Otkrivanjem tih komponenata, moguće je promotriti frekvenciju pojave poruka upozorenja nastalih između svih kvarova uzrokovanih tim komponentama. Primjenom klaster analize, sve poruke upozorenja koje su zabilježene između dva ili više ista kvara je moguće svrstati u različite grupe. Grupe poruka upozorenja bi zapravo predstavljale kategorije važnosti poruka upozorenja povezane s kvarom određene komponente. Drugim riječima, grupe poruka upozorenja koje bi bile kategoriziranje kao važne poruke upozorenja bi implicirale frekventnije poruke upozorenja prije pojave kvara na određenoj komponenti. Na ovaj način bi se olakšao postupak identifikacije grešaka bržim otkrivanjem neispravne komponente nakon pojave određenih poruka upozorenja.

## LITERATURA

- [1] Díaz-Ramírez, J.; Huertas, J.I.; Trigos, F.: „Aircraft maintenance, routing, and crew scheduling planning for airlines with a single fleet and a single maintenance and crew base“, *Computers & Industrial Engineering*, Vol. 75, pp. 68–78., 2014.
- [2] International Air Transport Association: „Airline Maintenance Cost Executive Commentary—An exclusive benchmark analysis (FY2014 data)“, IATA Maintenance Cost Task Force, 2015.
- [3] Sahay, A.: „The business of maintaining aircraft“ Ch. 2 of *Leveraging Information Technology for Optimal Aircraft Maintenance, Repair and Overhaul (MRO)*, 1st Edition, pp. 15–230., Woodhead Publishing Limited, Cambridge, UK, 2012., ISBN: 978–0–85709–143–7
- [4] Papakostas, N.; Papachatzakis, P.; Xanthakis, V.; Mourtzis, D.; Chryssolouris, G.: „An approach to operational aircraft maintenance planning“, *Decision Support Systems*, Vol. 48, No. 4, pp. 604–612., 2010.
- [5] Pleumpirom, Y.; Amornsawadwatana, S.: „Multiobjective Optimization of Aircraft Maintenance in Thailand Using Goal Programming: A Decision-Support Model“, *Advances in Decision Sciences*, Vol. 2012, pp. 1–17., 2012.
- [6] Yang, Z.; Yang, G.: „Optimization of Aircraft Maintenance plan based on Genetic Algorithm“, *Physics Procedia*, Vol. 33, pp. 580–586., 2012.
- [7] Regattieri, A.; Giazzi, A.; Gamberi, M.; Gamberini, R.: „An innovative method to optimize the maintenance policies in an aircraft: General framework and case study“, *Journal of Air Transport Management*, Vol. 44-45, pp. 8-20., 2015.
- [8] Qiang, F.; Songjie, L.; Bo, S.: „A multi-agent based intelligent configuration method for aircraft fleet maintenance personnel“, *Chinese Journal of Aeronautics*, Vol. 27, No. 2, pp. 280–290., 2014.
- [9] Belien, J., Demeulemeester, E.; De Bruecker, P.; Van den Bergh, J.; Cardoen, B.: „Integrated staffing and scheduling for an aircraft line maintenance problem“, *Computers & Operations Research*, Vol. 40, No. 4 , pp. 1023-1033., 2013.

- [10] Gavranis, A.; Kozanidis, G.: „An exact solution algorithm for maximizing the fleet availability of a unit of aircraft subject to flight and maintenance requirements“, *European Journal of Operational Research*, Vol. 242, No. 2, pp. 631–643., 2015.
- [11] Lapp, M.; Cohn, A.: „Modifying lines-of-flight in the planning process for improved maintenance robustness“, *Computers & Operations Research*, Vol. 39, No. 9, pp. 2051–2062., 2012.
- [12] Baohui, J., Chunhui, X., Yaohua, L.: „Study on Optimization Method of Aircraft Maintenance Plan Based on Longest Path“, *Journal of Applied Sciences*, Vol. 13, No. 16, pp. 3354–3357., 2013.
- [13] Başdere, M.; Bilge, Ü.: „Operational aircraft maintenance routing problem with remaining time consideration“, *European Journal of Operational Research*, Vol. 235, No. 1, pp. 315–328., 2014.
- [14] Van Horenbeek, A.: „Information-based maintenance optimization with focus on predictive maintenance“, doktorska disertacija, KU Leuven, Belgium, 2012.
- [15] Potes Ruiz, P.; Kamsu Foguem, B.; Grabot, B.: „Generating knowledge in maintenance from Experience Feedback“, *Knowledge-Based Systems*, Vol. 68, pp. 4–20., 2014.
- [16] Rodrigues, R. S.; Balestrassi, P. P.; Paiva, A. P.; Garcia-Diaz, A.; Pontes, F. J.: „Aircraft interior failure pattern recognition utilizing text mining and neural networks“, *Journal of Intelligent Information Systems*, Vol. 38, No. 3, pp. 741–766., 2012.
- [17] Gürbüz, F.; Özbakir, L.; Yapici, H.: „Classification rule discovery for the aviation incidents resulted in fatality“, *Knowledge-Based Systems*, Vol. 22, No. 8, pp. 622–632., 2009.
- [18] Arockia Christopher, A. B. ; Appavu alias Balamurugan, S.: „Prediction of warning level in aircraft accidents using data mining techniques“, *The Aeronautical Journal*, Vol. 118, No. 1206, pp. 935–952., 2014.
- [19] Bect, P.; Simeu-Abazi, Z.; Maisonneuve, P.-L.: „Identification of abnormal events by data monitoring: Application to complex systems“, *Computers in Industry*, Vol. 68, pp. 78–88., 2015.

- [20] Côme, E.; Cottrell, M.; Verleysen, M.; Lacaille, J.: „Aircraft engine health monitoring using Self-Organizing Maps“, *Advances in Data Mining. Applications and Theoretical Aspects. 10th Industrial Conference, ICDM 2010, Berlin, Germany, July 12-14*, pp. 405-417., 2010.
- [21] García Nieto, P. J.; García-Gonzalo, E.; Sánchez Lasheras, F.; de Cos Juez, F.J.: „Hybrid PSO–SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability“, *Reliability Engineering & System Safety*, Vol. 138, pp. 219–231., 2015.
- [22] Lv, Z.; Wang, J.; Zhang, G.: „Prognostics Health Management of Condition-Based Maintenance for Aircraft Engine Systems“. *2015 IEEE Conference on Prognostics and Health Management (PHM)*. Austin, TX, USA, June 22-25, 2015.
- [23] Wang, Z.-F.; Zarader, J.-L.; Argentieri, S.: „A Novel Aircraft Engine Fault Diagnostic and Prognostic System based on SVM“. *2012 IEEE International Conference on Condition Monitoring and Diagnosis*. Bali, Indonesia, Sept. 23-27, 2012.
- [24] Chen, C.; Brown, D.; Sconyers, C.; Zhang, B.; Vachtsevanos, G.; Orchard, M. E.: „An integrated architecture for fault diagnosis and failure prognosis of complex engineering systems“, *Expert Systems with Applications*, Vol. 39, No. 10, pp. 9031–9040., 2012.
- [25] Ferreira, S.; Arnaiz, A.; Sierra, B.; Irigoien, I.: „Application of Bayesian networks in prognostics for a new Integrated Vehicle Health Management concept“, *Expert Systems with Applications*, Vol. 39, No. 7, pp. 6402–6418., 2012.
- [26] Du, J.; Wang, S.; Han, L.; Zhao, S.; Guo, C.: „Prognostic management verification system of aircraft hydraulic power supply system“, *IEEE 10th International Conference on Industrial Informatics*. Beijing, China, July 25-27, 2012.
- [27] Rodrigues, L. R.; Gomes, J. P. P.; Ferri, F. A. S.; Medeiros, I. P.; Galvao, R. K. H.; Nascimento Junior, C. L.: „Use of PHM Information and System Architecture for Optimized Aircraft Maintenance Planning“, *IEEE Systems Journal*, Vol. 9, No. 4, pp. 1197–1207., 2015.
- [28] Viniacourt, F.; Woon, Y.-K.; Chan, K.; Ghelam, S.; Morel, H.; Chuan Tee, S.: „Operational data Processing and Techniques for Innovative MAINTenance (OPTIMAINT)“, *2011 Prognostics & System Health Management Conference, PHM2011, Shenzhen, China*, 2011.

- [29] Desforges, X.; Diévar, M.; Archimède, B.: „A prognostic function for complex systems to support production and maintenance co-operative planning based on an extension of object oriented Bayesian networks“, *Computers in Industry*, Vol. 86, pp. 34–51., 2017.
- [30] Baptista, M.; de Medeiros, I. P.; Malere, J. P.; Nascimento, C.; Prendinger, H.; Henriques, E. M.P.: „Comparative case study of life usage and data-driven prognostics techniques using aircraft fault messages“, *Computers in Industry*, Vol. 86, pp. 1–14., 2017.
- [31] Han, H. K.; Kim, H. S.; Sohn, S. Y.: „Sequential association rules for forecasting failure patterns of aircrafts in Korean airforce“, *Expert Systems with Applications*, Vol. 36, No. 2, pp. 1129–1133., 2009.
- [32] Julka, N.; Thirunavukkarasu, A.; Lendermann, P.; Gan, B. P.; Schirrmann, A.; Fromm, H.; Wong, E.: „Making use of prognostics health management information for aerospace spare components logistics network optimisation“, *Computers in Industry*, Vol. 62, No. 6, pp. 613–622., 2011.
- [33] Prescott, D.; Andrews, J.: „Modelling the Use of Maintenance to Minimise Aircraft Service Disruption“, *IFAC Proceedings Volumes*, Vol. 43, No. 3, pp. 44–49., 2010.
- [34] Jović, A.: „Dubinska analiza biomedicinskih vremenskih nizova zasnovana na računalnom radnom okviru za izlučivanje značajki“, doktorska disertacija, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, Zagreb, 2012.
- [35] NATO ARMP-7, Ed. 2: „Allied Reliability and Maintainability Publication“: NATO R&M terminology applicable to ARMPs, 2008.
- [36] C. Andela: „Improvements of Aircraft Availability within the Royal Netherlands Air Force“, RTO AVT-144 Workshop, Lithuania, October 3-5, 2006.
- [37] Fayyad, U. M.; Shapiro, G. P.; Smyth, P.: „The KDD process for extracting useful knowledge from volumes from data“. *Communication of ACM*, Vol. 39, No. 11, pp. 27–34, 1996.
- [38] Kumar, P.; Kumar Sehgal, V.; Chauhan, D.S.: „A Benchmark To Select Data Mining Based Classification Algorithms For Business Intelligence And Decision Support Systems“. *International Journal of Data Mining & Knowledge Management Process (IJDKP)* Vol.2, No.5, pp. 25–42., 2012.

- [39] Alwi, H.; Edwards, C.; Tan, C.P.: „Fault Tolerant Control and Fault Detection and Isolation“. In: *Fault Detection and Fault-Tolerant Control Using Sliding Modes. Advances in Industrial Control*, Springer-Verlag, London, 2011.
- [40] Carroll, J. S.; Johnson, E. J.: „Decision research: A field guide“. Newbury Park: Sage Publications. Chicago, 1990.
- [41] Lei, Z.: „Fault Prognostic Algorithm Based on Multivariate Relevance Vector Machine and Time Series Iterative Prediction“. In *Procedia Engineering*, Vol. 29, pp. 678–686., 2012.
- [42] Hu, C.: „Aircraft Fault Diagnosis Prognostics and Health Management Based on Flight Recorder“, 2012 Prognostics & System Health Management Conference (PHM-2012 Beijing), Beijing, May 23-25, 2012.
- [43] Tong, Z.; Guo, C.; Li, Y.: „Application Research on the PHM Technology in the Advanced Fight Plane Maintain and Guarantee“, *Aviation Maintenance & Engineering*, Vol. 1, pp. 37-39., 2009.
- [44] Elattar, H. M.; Elminir, H. K.; Riad, A. M.: „Prognostics: a literature review“, *Complex Intell. Syst.* Vol. 2, No. 2, pp. 125–154., 2016.
- [45] Yousuf, W. B.; Khan, T.; Ali, T.: „Prognostic Algorithms for Flaw Growth Prediction in an Aircraft Wing“. In *IEEE Trans. Rel.*, Vol. 66, No. 2, pp. 478–486., 2017.
- [46] Xiongzi, C.; Jinsong, Y.; Diyin, T.; Yingxun, W.: „Remaining Useful Life Prognostic Estimation for Aircraft Subsystems or Components: A Review“. *The 10th International Conference on Electronic Measurement & Instruments*. Chengdu, China, Aug. 16-19, 2011.
- [47] Pecht, M.; Jaai, R.: „A prognostics and health management roadmap for information and electronics-rich systems“. In *Microelectronics Reliability*, Vol. 50, No. 3, pp. 317–323., 2010.
- [48] Skormin, V. A.; Gorodetski, V. I.; Popyack, L. J.: „Data Mining Technology for Failure Prognostic of Avionics“. *IEEE Transactions on Aerospace and Electronic Systems*, Vol.38, No. 2, pp. 388–403., 2002.
- [49] Fayyad, U. M., Piatetsky-Shapiro, G.; Smyth, P.: „From data mining to knowledge discovery: An overview“. *Advances in Knowledge Discovery and Data Mining*, Cambridge, MA: MIT Press, pp. 1-34., 1995.

- [50] Núñez-del-Prado, M.; Gómez, R.: „Learning data analytics through a Problem Based Learning course“, 2017 IEEE World Engineering Education Conference (EDUNINE), pp. 52-56., 2017.
- [51] Sjöblom, O.: „Data Mining in Promoting Flight Safety“, doktorska disertacija, University of Turku, Finland, ISBN: 978-951-29-6546-5, 2016.
- [52] Maimon, O.; Rokach, L.: „Clustering methods“, In Data mining and knowledge discovery handbook, Springer, Berlin. ISBN: 978-0-387-24435-8, pp. 321–352., 2005.
- [53] Domazet-Lošo, M.: „Usporedba postupaka dubinske analize primijenjenih nad biološkim podacima“, magistarski rad, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, Zagreb, 2006.
- [54] Miner, G.; Elder, J.; Nisbet, R.: „Handbook of Statistical Analysis and Data Mining Applications“, Chapter 2: „The algorithms in data mining and text mining, the organization of the three most common data mining tools, and selected specialized areas using data mining“, ISBN: 978-0-12-374765-5, 2009.
- [55] Bishop, C.: „Neural Networks for Pattern Recognition“, Oxford, UK: Oxford University Press, New York, NY, USA, ISBN: 0198538642, 1995.
- [56] Bermúdez, Y.S.: „How to select the right machine learning approach?“, master thesis, Linnaeus University, Sweden, 2013.
- [57] John, G. H.; Langley, P.: „Estimating Continuous Distributions in Bayesian Classifiers“. In UAI'95 Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, San Mateo: Morgan Kaufmann Publishers Inc., 1995.
- [58] Helge, L.; Nielsen, T. D.: „Classification using Hierarchical Naïve Bayes models“, Machine Learning, Vol. 63, No. 2, pp. 135-159., 2006.
- [59] Ujević, F.: „Postupci analize podataka u izgradnji profila korisnika usluga“, magistarski rad, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, Zagreb, 2004.
- [60] Singh, S.; Selvakumar, S.: „A Hybrid Feature Subset Selection by Combining Filters and Genetic Algorithm“, International Conference on Computing, Communication and Automation (ICCCA2015), pp. 283–289., 2015.

- [61] Kashyap, H.; Das, S.; Bhattacharjee, J.; Halder, R.; Goswami, S.: „Multi-objective Genetic Algorithm setup for Feature Subset Selection in Clustering“, 3rd International Conference on Recent Advances in Information Technology (RAIT), pp. 243–247., 2016.
- [62] Nam, L. N. H.; Quoc, H. B.: „A combined approach for filter feature selection in document classification“, IEEE 27th International Conference on Tools with Artificial Intelligence, Nov. 9-11, Vietri sul Mare, Italy, pp. 317-324., 2015.
- [63] Oreški, S.: „Hibridne tehnike kombinatorne optimizacije temeljene na genetskim algoritmima s primjenom na odabir atributa u ocjenjivanju kreditnog rizika građana“, doktorska disertacija, Sveučilište u Zagrebu, Fakultet organizacije i informatike, Varaždin, 2014.
- [64] Alelyani, S.; Tang, J.; Liu, H.: „Feature Selection for Clustering: A Review“. Data Clustering: Algorithms and Applications, Vol. 29, 2013.
- [65] Yin, L.; Ge, Y.; Xiao, J.; Wang, X.; Quan, X.: „Feature selection for high-dimensional imbalanced data“, Neurocomputing, Vol. 105, pp. 3–11., 2013.
- [66] Wong, T.-T.: „Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation“, Pattern Recognition, Vol. 48, No. 9, pp. 2839-2846., 2015.
- [67] Hu, C.; Youn, B. D.; Wang, P.; Taek Yoon, J.: „Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life“, In Reliability Engineering & System Safety, Vol. 103, pp. 120–135., 2012.
- [68] Bekkar, M.; Djemaa, H. K.; Alitouche, T. A.: „Evaluation Measures for Models Assessment over Imbalanced Data Sets“, Journal of Information Engineering and Applications, Vol.3, No.10, 2013.
- [69] Abd Elrahman, S. M.; Abraham, A.: „A Review of Class Imbalance Problem“, Journal of Network and Innovative Computing, Vol. 1, pp. 332-340., 2013.
- [70] Kumar, Y.; Sahoo, G.: „Analysis of Parametric & Non Parametric Classifiers for Classification Technique using Weka“, I.J. Information Technology and Computer Science, Vol. 7, pp. 43-49., 2012.
- [71] Reif, M.; Shafait, F.; Dengel, A.: „Meta-learning for evolutionary parameter optimization of classifiers“. Machine Learning, Vol. 87, No. 2, pp. 357–380., 2012.



- [72] Ma, X.; Zhang, Y.; Wang, Y.: „Performance evaluation of kernel functions based on grid search for Support Vector Regression“, Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), July 15-17, Siem Reap, Cambodia, 2015.
- [73] Naik, A.; Samant, L.: „Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime“. *Procedia Computer Science*, pp. 662–668., 2016.
- [74] Chen, X.; Williamsand, G.; Xu, X.: „A Survey of Open Source Data Mining Systems“, In *Emerging Technologies in Knowledge Discovery and Data Mining*, Springer Berlin Heidelberg, pp. 3-14., 2007.
- [75] Al-Odan, H. A.; Al-Daraiseh, A. A.: „Open Source Data Mining Tools: A Comparative Study“, 1st International Conference on Electrical and Information Technologies (ICEIT'2015), March 25-27, pp. 369-374., 2015.
- [76] Gartner, Inc.: „Magic Quadrant for Data Science Platforms“, by Alexander Linden, Peter Krensky, Jim Hare, Carlie J. Idoine, Svetlana Sicular, Shubhangi Vashisth, February 14th, 2017.
- [77] Schmidt, B.; Sandberg, U.; Wang, L.: „Next Generation Condition Based Predictive Maintenance“, *Proceedings of the 6th International Swedish Production Symposium*, September 16-18, 2014.
- [78] Zarei, J.; Tajeddini, M. A.; Karimi, H. R.: „Vibration analysis for bearing fault detection and classification using an intelligent filter“. *Mechatronics*. Vol. 24, No. 2, pp.151-257., 2014.
- [79] Paliwal, D.; Choudhury, A.; Govardhan, T.; Chandrawat, S. S.: „A Novel Approach for Detection of Bearing Defects from Noisy Vibration Signal“. *Applied Mechanics and Materials: Trans Tech Publ*; Vol. 592-594, pp. 2001-2005., 2014.
- [80] Prieto, M. D.; Cirrincione, G.; Espinosa, A. G.; Ortega, J. A.; Henao, H.: „Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks“. *IEEE Transactions on Industrial Electronics*. Vol. 60, No. 8, pp. 3398-3407., 2013.

- [81] Raišutis, R.; Jasiūnienė, E.; Šliteris, R.; Vladiškauskas, A.: „The review of non-destructive testing techniques suitable for inspection of the wind turbine blades“. *Ultragarsas Journal*, Vol. 63, No. 1, pp. 49-52., 2016.
- [82] Shanmukha Priya, V.; Ramesh, M. R.; Naidu, V. P. S.: „Bearing Health Condition Monitoring: Frequency domain analysis“. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, Vol. 3, No. 5, pp. 260-268., 2014.
- [83] Goel, S.; Ghosh, R.; Kumar, S.; Akula, A. A.: „Methodical Review of Condition Monitoring Techniques For Electrical Equipment“. *National Seminar & Exhibition on Non-Destructive Evaluation (NDE 2014)*, Pune, 2014.
- [84] Stenström, C.; Carlson, J.; Lundberg, J.: „Condition monitoring of cracks and wear in mining mills using water squirter ultrasonics“. *International Journal of Condition Monitoring*, Vol. 5, pp. 2-8., 2015.
- [85] Bernardo, J. T.: „Application of the JDL data fusion process model to hard/soft information fusion in the condition monitoring of aircraft“. *Proceedings of SPIE*, Vol. 9122, 2014.
- [86] Liu, L.; Wang, S.; Liu, D.; Zhang, Y.; Peng, Y.: „Entropy-based sensor selection for condition monitoring and prognostics of aircraft engine“. *Microelectronics Reliability*. Vol. 55, No. 9-10, pp. 2092-2096., 2015.
- [87] Hall, D. L.; McMullen, S. A.; Hall, C. M.: „New perspectives on level-5 information fusion: The impact of advances in information technology and user behavior“. *2015 IEEE International Conference on: Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pp. 214-219., 2015.
- [88] Fang, X.; Gebraeel, N.; Paynabar, K.: „Scalable Prognostic Models for Large-scale Condition Monitoring Applications“. *Qualiti & Reliability Engineering*, pp. 698-710., 2016.
- [89] Xu, J.; Wang, Y.; Xu, L.: „PHM-oriented integrated fusion prognostics for aircraft engines based on sensor data“. *IEEE Sensors Journal*, Vol. 14, No. 4, pp. 1124-1132., 2014.
- [90] Sai Kiran, P. V. R.; Adithya, R.; Balasubramanian, N.; Upendranath, V.: „A Novel Data Fusion Method for Incipient Fault Detection in TRU of Aircraft Electrical System“. In: *ISSS International Conference on Smart Materials, Structures and Systems*, July 08-11, Bangalore, India, 2014.

- [91] Liu, K.; Huang, S.: „Integration of data fusion methodology and degradation modeling process to improve prognostics“. IEEE Transactions on Automation Science and Engineering, Vol. 13, No. 1, pp. 344-354., 2016.
- [92] Liu, Z.; Mrad, N.: „Data fusion for the diagnostics, prognostics, and health management of aircraft systems“. Foundations and Practical Applications of Cognitive Systems and Information Processing: Springer; pp. 389-399., 2014.
- [93] Gerdes, M.; Galar, D.; Scholz, D.: „Automated parameter optimization for feature extraction for condition monitoring“. 14th IMEKO TC10 Workshop on Technical Diagnostics 2016: New Perspectives in Measurements, Tools and Techniques for Systems Reliability, Maintainability and Safety, Milan, Italy, June 27-28, pp. 452-457., 2016.
- [94] Air Transport Association of America: „ATA Specification 100–Specification for Manufacturers' Technical Data“, Washington DC, USA, Revision No. 37, 1999.
- [95] Airbus: „Single aisle technical training manual T1+T2 (CFM 56) (Lv1 2&3) INDICATING“, 2008.
- [96] Sutaria, R. D.: „Compliant Aviation Reliability Reporting“, Avia Intelligence, Dubai, 2016.
- [97] Amborski, J.: „Calculation of Alert Levels for Reliability“. Paper presented at 23rd European Conference on Modelling and Simulation. Madrid, Spain, 2009.
- [98] RapidMiner GmbH.: „RapidMiner 7 Operator Reference Manual“, 2016.
- [99] Srivastava, A.N.; Zane-Ulman, B.: „Discovering recurring anomalies in text reports regarding complex space systems“, Aerospace Conference, 2005 IEEE, March 5-12, Big Sky, MT, USA, 2005.
- [100] Grljević, O.: „Sentiment u sadržajima sa društvenih mreža kao instrument unapređenja poslovanja visokoškolskih institucija“, doktorska disertacija, Sveučilište u Novom Sadu, Subotica, 2016.
- [101] Tian, X.; Xanmei, C.: „An Improvement to TF-IDF: Term Distribution based Term Weight Algorithm“, Journal of Software, Vol. 6, No. 3, pp. 413-420., 2011.

- [102] Verma, T.; Renu, Gaur, D.: „Tokenization and Filtering Process in RapidMiner“, International Journal of Applied Information Systems, Vol. 7, No. 2, pp. 16-18., 2014.
- [103] Lee, I.-H.; Lushington, G. H.; Visvanathan, M.: „A filter-based feature selection approach for identifying potential biomarkers for lung cancer“, Journal of Clinical Bioinformatics, Vol. 1, pp. 1-8., 2011.
- [104] Moir, I.; Seabridge, A.; Jukes, M.: „Civil Avionics Systems“, 2nd Edition, John Wiley & Sons, 2013., ISBN: 978-1-118-34180-3
- [105] Wtruib training: „Airbus maintenance training“, <http://wtruib.ru>, datum pristupa: 13.08.2017.

# PRILOZI

## Prilog A – Programska izvedba algoritma za izdvajanje poruka upozorenja

```
Private Sub UCITAJ_PODATKE_1_Click()
    Dim msgbox_odgovor As String
    Dim TextLine
    Dim pocetna_sekvenca As String
    Dim baza As Database
    Dim rs As Recordset
    Dim linija As Double
    DoCmd.RunSQL "DELETE T_Zrakoplov1.* FROM T_Zrakoplov1;"
    baza = DBEngine.Workspaces(0).Databases(0)
    rs = baza.OpenRecordset("T_ZRAKOPLOV1", DB_OPEN_DYNASET)
    Open "C:\A_Marina\ZRAKOPLOV1.txt" For Input As #1
    linija = 0
    Do While Not EOF(1)
        Line Input #1, TextLine
        linija = linija + 1
        Me!TXT_LINIJA = linija
        pocetna_sekvenca = Trim(Left(TextLine, 7))
        If pocetna_sekvenca = "- MPF/" Then
            Dim registracija As String
            registracija = Mid(Trim(TextLine), 11, 6)
            Dim broj_leta As String
            broj_leta = Mid(Trim(TextLine), 20, 6)
            Dim godina As String
            godina = Mid(Trim(TextLine), 33, 2)
            Dim mjesec As String
            mjesec = Mid(Trim(TextLine), 35, 2)
            Dim dan As String
            dan = Mid(Trim(TextLine), 37, 2)
            Dim DATUM As String
            If IsNumeric(godina) And IsNumeric(mjesec) And IsNumeric(dan) Then
                DATUM = dan + "." + mjesec + "." + godina
            Else
                DATUM = "00" + "." + "00" + "." + "00"
            End If
            Dim sat As String
            sat = Mid(Trim(TextLine), 39, 2)
            Dim minuta As String
            minuta = Mid(Trim(TextLine), 41, 2)
            Dim VRIJEME As String
            If IsNumeric(sat) And IsNumeric(minuta) Then
                VRIJEME = sat + ":" + minuta
            Else
                VRIJEME = "00" + ":" + "00"
            End If
            Dim odlazni_aerodrom As String
            odlazni_aerodrom = Mid(Trim(TextLine), 48, 4)
            Dim dolazni_aerodrom As String
            dolazni_aerodrom = Mid(Trim(TextLine), 55, 4)
            Dim warning_poruka As String
            warning_poruka = Mid(Trim(TextLine), 60, 2)
            If warning_poruka = "WN" Then
                godina = Mid(Trim(TextLine), 62, 2)
                mjesec = Mid(Trim(TextLine), 64, 2)
                dan = Mid(Trim(TextLine), 66, 2)
                Dim date_warning_poruka As String
                If IsNumeric(dan) And IsNumeric(mjesec) And IsNumeric(godina) Then
```

```

    date_warning_poruka = dan + "." + mjesec + "." + godina
Else
    date_warning_poruka = "00" + "." + "00" + "." + "00"
End If
sat = Mid(Trim(TextLine), 68, 2)
minuta = Mid(Trim(TextLine), 70, 2)
Dim time_warning_poruka As String
If IsNumeric(sat) And IsNumeric(minuta) Then
    time_warning_poruka = sat + ":" + minuta
Else
    time_warning_poruka = "00" + ":" + "00"
End If
Dim ata As String
ata = Mid(Trim(TextLine), 74, 2)
Dim pod_ata As String
pod_ata = Mid(Trim(TextLine), 76, 2)
Dim pod_pod_ata As String
pod_pod_ata = Mid(Trim(TextLine), 78, 2)
Dim faza_leta As String
faza_leta = Mid(Trim(TextLine), 80, 2)
Dim zabiljezana_poruka As String
zabiljezana_poruka = Mid(Trim(TextLine), 82, 1000)
rs.AddNew()
rs![registracija] = registracija
rs![broj_leta] = broj_leta
rs![DATUM] = DATUM
rs![VRIJEME] = VRIJEME
rs![odlazni_aerodrom] = odlazni_aerodrom
rs![dolazni_aerodrom] = dolazni_aerodrom
rs![warning_poruka] = warning_poruka
If IsNumeric(dan) And IsNumeric(mjesec) And IsNumeric(godina) Then
    rs![date_warning_poruka] = date_warning_poruka
Else
    rs![date_warning_poruka] = "00" + "." + "00" + "." + "00"
End If
If IsNumeric(sat) And IsNumeric(minuta) Then
    rs![time_warning_poruka] = time_warning_poruka
Else
    rs![time_warning_poruka] = "00" + ":" + "00" + ":" + "00"
End If
rs![ata] = ata
rs![pod_ata] = pod_ata
rs![pod_pod_ata] = pod_pod_ata
rs![faza_leta] = faza_leta
rs![zabiljezana_poruka] = zabiljezana_poruka
rs.Update()
End If
Line Input #1, TextLine
linija = linija + 1
Me!TXT_LINIJA = linija
End If
Do While pocetna_sekvenca = "- MPF/" And Trim(Left(TextLine, 1)) = ","
    If warning_poruka = "WN" And Len(Mid(Trim(TextLine), 2, 100000)) > 0 Then
        godina = Mid(Trim(TextLine), 2, 2)
        mjesec = Mid(Trim(TextLine), 4, 2)
        dan = Mid(Trim(TextLine), 6, 2)
        Dim datum_warning_poruka As String
        Dim date_warning_poruka1 As String
        If IsNumeric(dan) And IsNumeric(mjesec) And IsNumeric(godina) Then
            datum_warning_poruka = dan + "." + mjesec + "." + godina
            date_warning_poruka1 = datum_warning_poruka
        Else
            datum_warning_poruka = "00" + "." + "00" + "." + "00"
            date_warning_poruka1 = datum_warning_poruka
        End If
        sat = Mid(Trim(TextLine), 8, 2)
    End If

```

```

minuta = Mid(Trim(TextLine), 10, 2)
Dim vrijeme_warning_poruka As String
Dim time_warning_poruka1 As String
If IsNumeric(sat) And IsNumeric(minuta) Then
    vrijeme_warning_poruka = sat + ":" + minuta
    time_warning_poruka1 = vrijeme_warning_poruka
Else
    vrijeme_warning_poruka = "00" + ":" + "00"
    time_warning_poruka1 = vrijeme_warning_poruka
End If
ata = Mid(Trim(TextLine), 14, 2)
pod_ata = Mid(Trim(TextLine), 16, 2)
pod_pod_ata = Mid(Trim(TextLine), 18, 2)
faza_leta = Mid(Trim(TextLine), 20, 2)
zabiljezana_poruka = Mid(Trim(TextLine), 22, 1000)
rs.AddNew()
rs![registracija] = registracija
rs![broj_leta] = broj_leta
rs![DATUM] = DATUM
rs![VRIJEME] = VRIJEME
rs![odlazni_aerodrom] = odlazni_aerodrom
rs![dolazni_aerodrom] = dolazni_aerodrom
rs![warning_poruka] = warning_poruka
If IsNumeric(dan) And IsNumeric(mjesec) And IsNumeric(godina) Then
    rs![date_warning_poruka] = date_warning_poruka1
Else
    rs![date_warning_poruka] = "00" + "." + "00" + "." + "00"
End If
If IsNumeric(sat) And IsNumeric(minuta) Then
    rs![time_warning_poruka] = time_warning_poruka1
Else
    rs![time_warning_poruka] = "00" + ":" + "00"
End If
rs![ata] = ata
rs![pod_ata] = pod_ata
rs![pod_pod_ata] = pod_pod_ata
rs![faza_leta] = faza_leta
rs![zabiljezana_poruka] = zabiljezana_poruka
rs.Update()
End If
Line Input #1, TextLine
linija = linija + 1
Me!TXT_LINIJA = linija
Loop
Loop
Close #1 'Zatvaranje datoteke ZRAKOPLOV1.txt
rs.Close()
msgbox_odgovor = MsgBox("PODACI ZRAKOPLOV UCITANI", vbInformation, "FSB-IIM")
End Sub

```

## Prilog B – Programska izvedba algoritma za otkrivanje uzoraka iz skupa zapisa o prošlim nepravilnostima/kvarovima nastalih na temelju uzoraka iz skupa zapisa o porukama upozorenja

```
Function ProvjeriPostojanjeDatumDog(datum, dogadjaj)
    Dim rangIspod As Range, vel As Integer, velIspod As Integer, datum2 As Date, ima As Boolean
    With Sheets("Zapis o porukama upozorenja")
        stavkeSkup = .Range(.Cells(3, 1), .Cells(3, 1).End(xlDown))
    End With
    brojRedova = stavkeSkup.Rows.Count
    For i = 3 To brojRedova + 2
        datum2 = Worksheets("Zapis o porukama upozorenja").Cells(i, 1).Value
        Dim c As Range, trazi As String, kolona As Integer, dogPostoji As Integer, rez As Integer
        trazi = "DOG " & dogadjaj & "_Ukupno"
        If datum = datum2 Then
            With Worksheets(2).Range("B2:CZ2")
                c = .Find(what:=trazi, LookIn:=xlValues)
                rez = WorksheetFunction.CountIf(Worksheets(2).Range("B2:CZ2"), trazi)
                If WorksheetFunction.CountIf(Worksheets(2).Range("B2:CZ2"), trazi) > 0 Then
                    kolona = c.Column
                    dogPostoji = Worksheets("Zapis o porukama upozorenja").Cells(i, kolona).Value
                Else
                    dogPostoji = 0
                End If
            End With
            If dogPostoji > 0 Then
                ima = True
                Exit For
            End If
        End With
    Next i
    If ima = True Then
        ProvjeriPostojanjeDatumDog = ima
    End If
End Function
```

---

```
Function ProvjeriPostojanjeDatum(datum, indeks)
    Dim rangIspod As Range, vel As Integer, velIspod As Integer, datum2 As Date, ima As Boolean, dogadjaj2 As Integer
    With Sheets("Zapis o kvarovima")
        stavkeSkup = .Range(.Cells(2, 1), .Cells(2, 1).End(xlDown))
    End With
    brojRedova = stavkeSkup.Rows.Count
    For i = 2 To brojRedova + 1
        datum2 = Worksheets("Zapis o kvarovima").Cells(i, 1).Value
        dogadjaj2 = Worksheets("Zapis o kvarovima").Cells(i, 2).Value
        Dim c As Range, trazi As String, kolona As Integer, dogPostoji As Integer
        trazi = "DOG " & dogadjaj2 & "_Ukupno"
        If datum = datum2 Then
            With Worksheets(2).Range("B2:CZ2")
                c = .Find(what:=trazi, LookIn:=xlValues)
                If WorksheetFunction.CountIf(Worksheets(2).Range("B2:CZ2"), trazi) > 0 Then
                    kolona = c.Column
                    dogPostoji = Worksheets("Zapis o porukama upozorenja").Cells(indeks, kolona).Value
                Else
                    dogPostoji = 0
                End If
            End With
            If dogPostoji > 0 Then
```



```

        ima = True
    Exit For
End If
End With
End If
Next i
If ima = True Then
    ProvjeriPostojanjeDatum = ima
End If
End Function

```

---

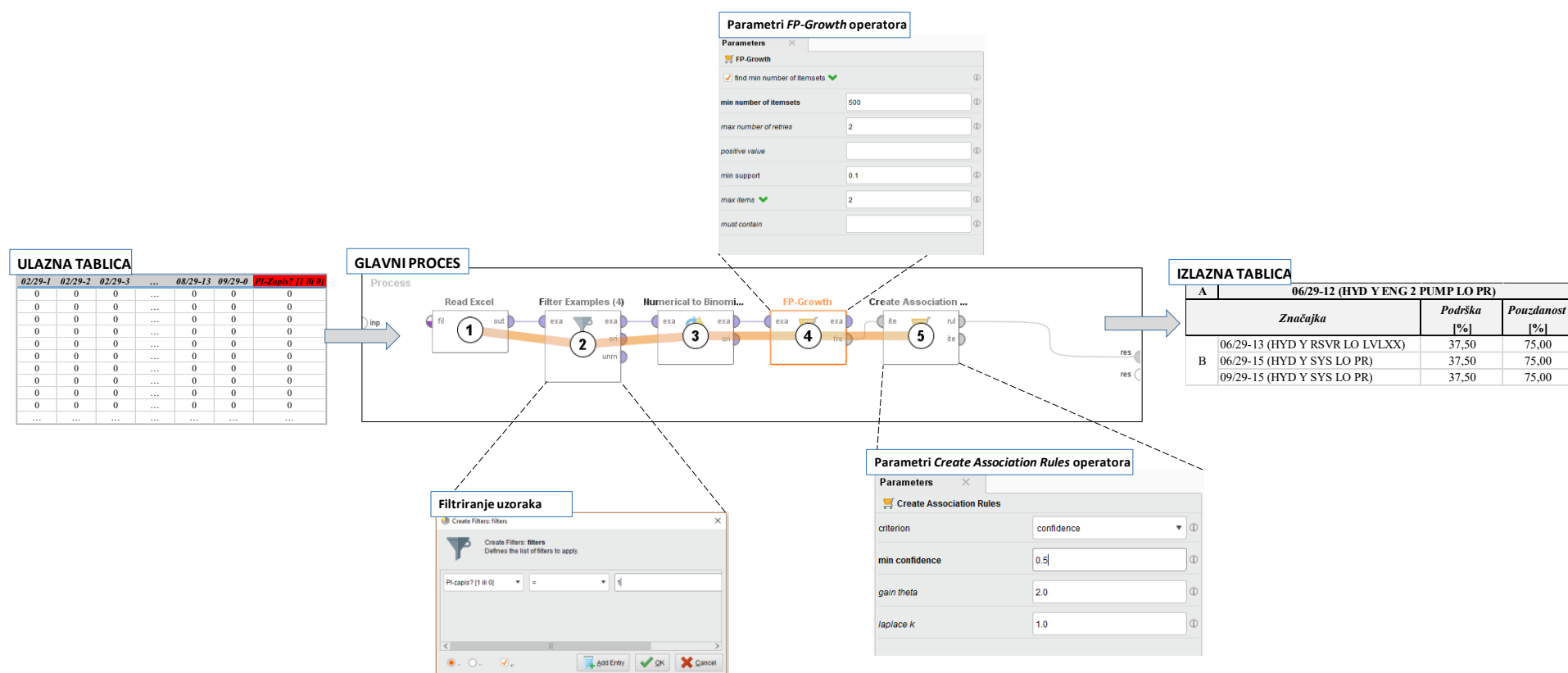
```

Function DohvatiDog(datum, indeks)
    Dim rangIspod As Range, vel As Integer, velIspod As Integer, datum2 As Date, ima As Boolean, dogadjaj2 As Integer
    With Sheets("Zapis o kvarovima")
        stavkeSkup = .Range(.Cells(2, 1), .Cells(2, 1).End(xlDown))
    End With
    brojRedova = stavkeSkup.Rows.Count
    For i = 2 To brojRedova + 1
        datum2 = Worksheets("Zapis o kvarovima").Cells(i, 1).Value
        dogadjaj2 = Worksheets("Zapis o kvarovima").Cells(i, 2).Value
        Dim c As Range, trazi As String, kolona As Integer, dogPostoji As Integer
        trazi = "DOG " & dogadjaj2 & "_Ukupno"
        If datum = datum2 Then
            dogPostoji = Worksheets("Zapis o porukama upozorenja").Cells(indeks, 106).Value
            If dogPostoji = 1 Then
                ima = True
            Exit For
            End If
        End If
    Next i
    If ima = True Then
        DohvatiDog = dogadjaj2
    Else
        DohvatiDog = 0
    End If
End Function

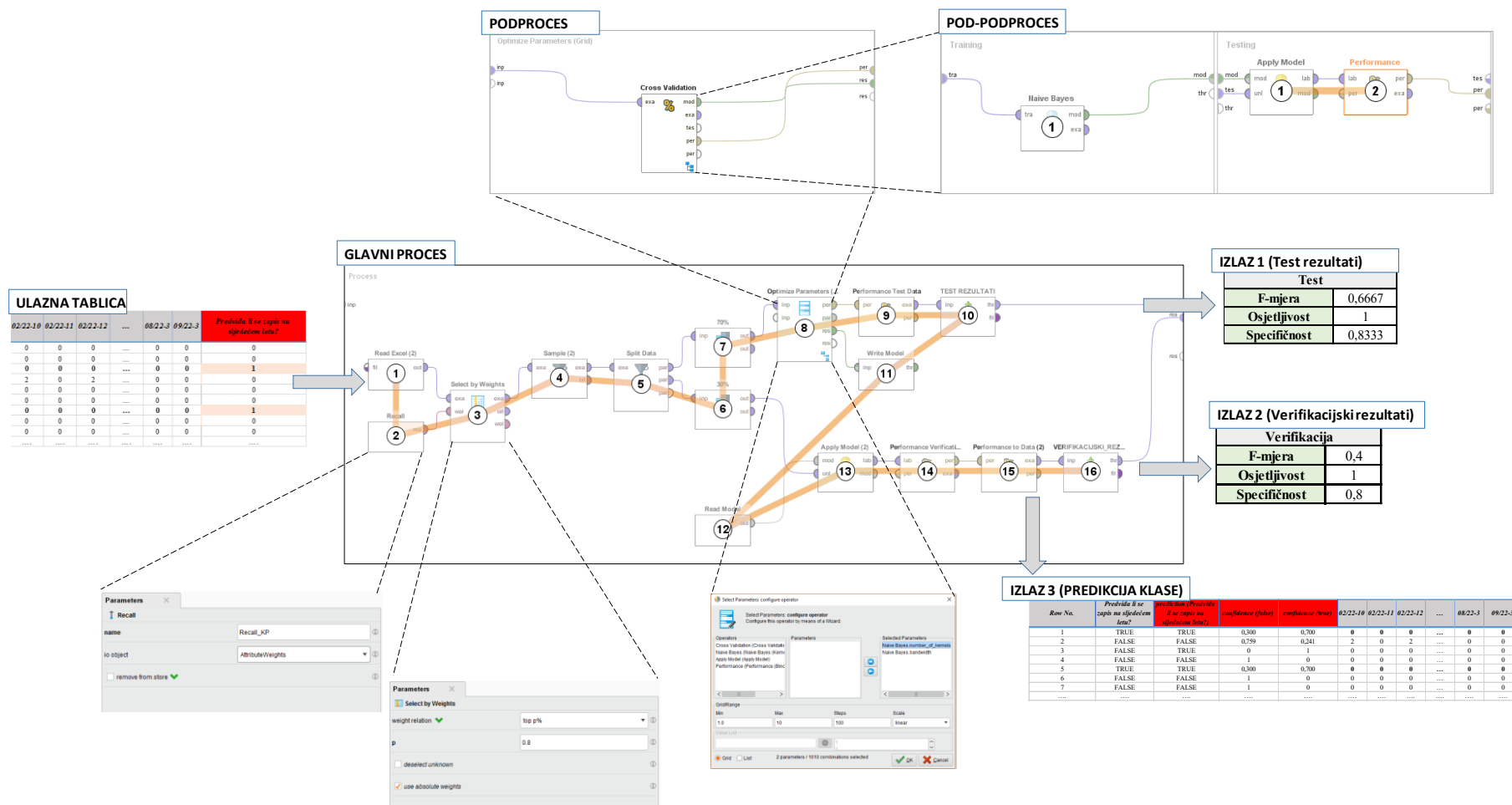
```



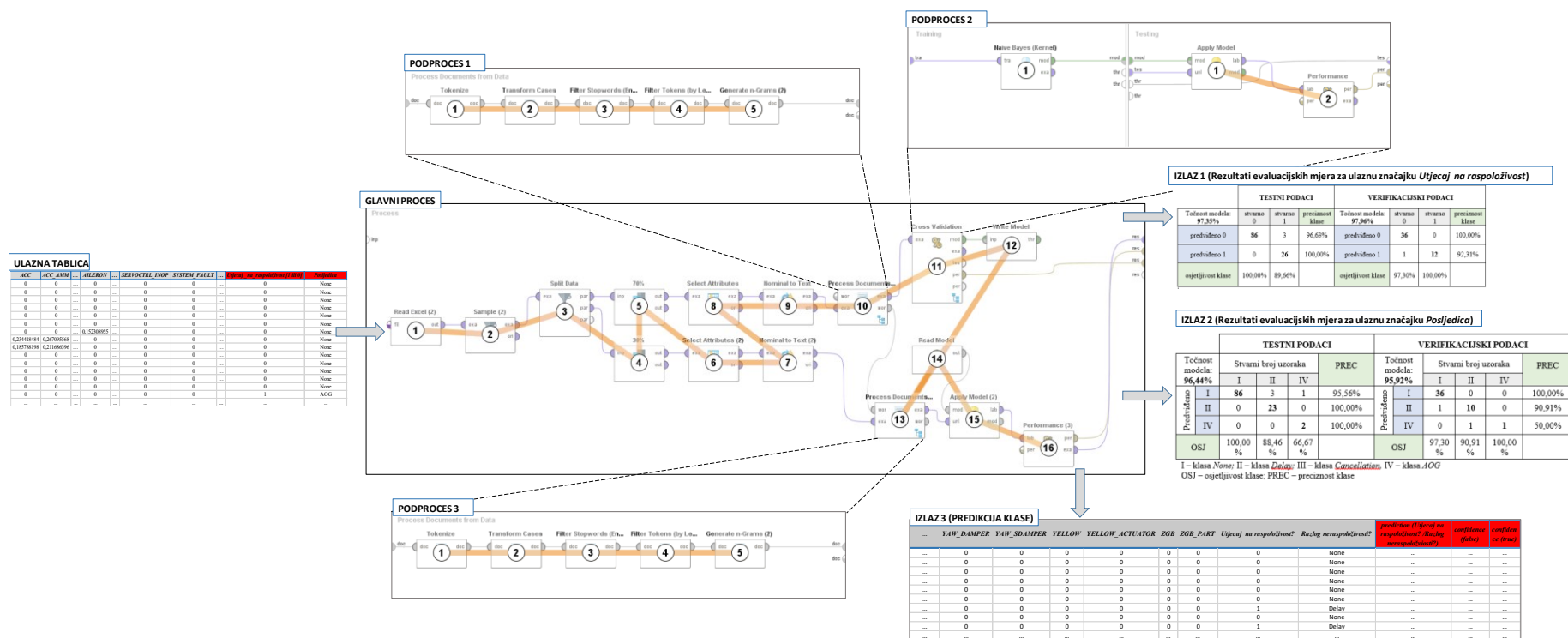
## Prilog C.2. Proces otkrivanja asocijativnih pravila



## Prilog C.3. Proces izgradnje klasifikacijskih modela prve grupe



## Prilog C.4. Proces izgradnje klasifikacijskih modela druge grupe



## Prilog D – Rang značajki analiziranih sustava dobiven filterskim postupcima za selekciju značajki

Prilog D.1. Rang značajki za sustav ATA 22

ATA 22 – rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
1.	06/22-2	06	22-2	AUTO FLT A/THR OFF	0,4038	0,0391	0,0855	0,5438	1,0000
2.	06/22-14	06	22-14	AUTO FLT YAW DAMPER 1	0,3629	0,0206	0,0453	0,2879	0,6685
3.	04/22-12	04	22-12	AUTO FLT RUD TRV LIM 1	0,2476	0,0096	0,0205	0,4038	0,6356
4.	06/22-15	06	22-15	AUTO FLT YAW DAMPER 2	0,3151	0,0156	0,0367	0,1959	0,5252
5.	02/22-10	02	22-10	AUTO FLT RUD TRIM1 FAULT	0,0463	0,0096	0,0205	0,4038	0,4478
6.	02/22-14	02	22-14	AUTO FLT YAW DAMPER 1	0,0303	0,0096	0,0205	0,4038	0,4329
7.	09/22-13	09	22-13	AUTO FLT RUD TRV LIM SYS	0,2356	0,0135	0,0446	0,0774	0,3462
8.	03/22-2	03	22-2	AUTO FLT A/THR OFF	0,2212	0,0077	0,0255	0,0515	0,2853
9.	02/22-4	02	22-4	AUTO FLT FAC 1 FAULTXX	0,2060	0,0066	0,0196	0,0671	0,2792
10.	06/22-11	06	22-11	AUTO FLT RUD TRIM2 FAULT	0,2060	0,0066	0,0196	0,0671	0,2792
11.	03/22-13	03	22-13	AUTO FLT RUD TRV LIM SYS	0,2107	0,0070	0,0221	0,0549	0,2748
12.	02/22-6	02	22-6	AUTO FLT FCU 1X2 FAULT	0,1884	0,0056	0,0171	0,0541	0,2474
13.	02/22-7	02	22-7	AUTO FLT FCU 1 FAULT	0,1429	0,0066	0,0196	0,0671	0,2203
14.	02/22-2	02	22-2	AUTO FLT A/THR OFF	0,1359	0,0040	0,0131	0,0363	0,1765
15.	06/22-12	06	22-12	AUTO FLT RUD TRV LIM 1	0,1388	0,0036	0,0136	0,0300	0,1735
16.	02/22-11	02	22-11	AUTO FLT RUD TRIM2 FAULT	0,1011	0,0027	0,0218	0,0351	0,1497
17.	02/22-13	02	22-13	AUTO FLT RUD TRV LIM SYS	0,0971	0,0015	0,0124	0,0282	0,1298
18.	02/22-15	02	22-15	AUTO FLT YAW DAMPER 2	0,0895	0,0019	0,0157	0,0307	0,1286
19.	02/22-16	02	22-16	AUTO FLT YAW DAMPER SYS	0,0885	0,0014	0,0116	0,0275	0,1203
20.	07/22-2	07	22-2	AUTO FLT A/THR OFF	0,0867	0,0012	0,0100	0,0262	0,1157
21.	02/22-9	02	22-9	AUTO FLT RUD TRIM SYS	0,0831	0,0011	0,0092	0,0255	0,1109

ATA 22 – rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
22.	06/22-13	06	22-13	AUTO FLT RUD TRV LIM SYS	0,0805	0,0011	0,0092	0,0255	0,1084
23.	06/22-16	06	22-16	AUTO FLT YAW DAMPER SYS	0,0775	0,0010	0,0084	0,0248	0,1042
24.	06/22-9	06	22-9	AUTO FLT RUD TRIM SYS	0,0759	0,0010	0,0084	0,0248	0,1027
25.	09/22-1	09	22-1	AUTO FLT	0,0889	0,0012	0,0045	0,0134	0,1008
26.	02/22-12	02	22-12	AUTO FLT RUD TRV LIM 1	0,0659	0,0027	0,0156	0,0191	0,0963
27.	06/22-3	06	22-3	AUTO FLT AP OFF	0,0512	0,0026	0,0084	0,0314	0,0873
28.	07/22-3	07	22-3	AUTO FLT AP OFF	0,0625	0,0006	0,0053	0,0218	0,0841
29.	09/22-12	09	22-12	AUTO FLT RUD TRV LIM 1	0,0539	0,0007	0,0061	0,0226	0,0777
30.	02/22-5	02	22-5	AUTO FLT FAC 2 FTXX	0,0546	0,0005	0,0045	0,0210	0,0752
31.	04/22-2	04	22-2	AUTO FLT A/THR OFF	0,0525	0,0004	0,0038	0,0200	0,0716
32.	06/22-7	06	22-7	AUTO FLT FCU 1 FAULT	0,0460	0,0005	0,0045	0,0210	0,0671
33.	09/22-3	09	22-3	AUTO FLT AP OFF	0,0442	0,0003	0,0030	0,0190	0,0621
34.	03/22-12	03	22-12	AUTO FLT RUD TRV LIM 1	0,0545	0,0005	0,0017	0,0081	0,0604
35.	02/22-8	02	22-8	AUTO FLT FCU 2 FAULT	0,0404	0,0003	0,0022	0,0179	0,0567
36.	06/22-1	06	22-1	AUTO FLT	0,0404	0,0003	0,0022	0,0179	0,0567
37.	05/22-12	05	22-12	AUTO FLT RUD TRV LIM 1	0,0404	0,0003	0,0022	0,0179	0,0567
38.	02/22-3	02	22-3	AUTO FLT AP OFF	0,0404	0,0003	0,0022	0,0179	0,0567
39.	05/22-2	05	22-2	AUTO FLT A/THR OFF	0,0404	0,0003	0,0022	0,0179	0,0567
40.	03/22-1	03	22-1	AUTO FLT	0,0404	0,0003	0,0022	0,0179	0,0567
41.	06/22-0	06	22-0	A FLT A/THR OFFX	0,0404	0,0003	0,0022	0,0179	0,0567
42.	08/22-3	08	22-3	AUTO FLT AP OFF	0,0404	0,0003	0,0022	0,0179	0,0567
43.	06/22-8	06	22-8	AUTO FLT FCU 2 FAULT	0,0329	0,0002	0,0015	0,0165	0,0476
44.	07/22-12	07	22-12	AUTO FLT RUD TRV LIM 1	0,0329	0,0002	0,0015	0,0165	0,0476
45.	05/22-3	05	22-3	AUTO FLT AP OFF	0,0312	0,0002	0,0015	0,0165	0,0460
46.	06/22-10	06	22-10	AUTO FLT RUD TRIM1 FAULT	0,0050	0,0011	0,0084	0,0248	0,0366

ATA 22 – rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
47.	08/22-13	08	22-13	AUTO FLT RUD TRV LIM SYS	0,0232	0,0001	0,0007	0,0146	0,0360
48.	04/22-8	04	22-8	AUTO FLT FCU 2 FAULT	0,0232	0,0001	0,0007	0,0146	0,0360
49.	09/22-2	09	22-2	AUTO FLT A/THR OFF	0,0232	0,0001	0,0007	0,0146	0,0360
50.	02/22-1	02	22-1	AUTO FLT	0,0232	0,0001	0,0007	0,0146	0,0360
51.	08/22-12	08	22-12	AUTO FLT RUD TRV LIM 1	0,0232	0,0001	0,0007	0,0146	0,0360
52.	05/22-16	05	22-16	AUTO FLT YAW DAMPER SYS	0,0000	0,0000	0,0000	0,0000	0,0000
53.	08/22-16	08	22-16	AUTO FLT YAW DAMPER SYS	0,0000	0,0000	0,0000	0,0000	0,0000
54.	08/22-2	08	22-2	AUTO FLT A/THR OFF	0,0000	0,0000	0,0000	0,0000	0,0000



Prilog D.2. Rang značajki za sustav ATA 27

ATA 27- rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
1.	02/27-16	02	27-16	F/CTL L AIL FAULT	0,299	0,011	0,025	0,453	1,000
2.	02/27-01	02	27-01	F/CTL AIL SERVO FAULT	0,207	0,011	0,025	0,453	0,880
3.	06/27-27	06	27-27	F/CTL SPLR FAULT	0,211	0,006	0,013	0,403	0,795
5.	09/27-27	09	27-27	F/CTL SPLR FAULT	0,211	0,006	0,013	0,403	0,795
4.	02/27-09	02	27-09	F/CTL FCDC 1 FAULT	0,211	0,006	0,013	0,403	0,795
6.	06/27-06	06	27-06	F/CTL ELAC 2 FAULT	0,196	0,006	0,015	0,403	0,779
7.	02/27-06	02	27-06	F/CTL ELAC 2 FAULT	0,197	0,006	0,013	0,403	0,777
8.	02/27-21	02	27-21	F/CTL SEC 3 FAULT	0,176	0,006	0,013	0,403	0,749
9.	06/27-01	06	27-01	F/CTL AIL SERVO FAULT	0,143	0,006	0,013	0,403	0,705
10.	02/27-02	02	27-02	F/CTL ALTN LAW	0,172	0,007	0,017	0,215	0,505
11.	02/27-19	02	27-19	F/CTL SEC 1 FAULT	0,158	0,002	0,009	0,116	0,339
12.	07/27-28	07	27-28	F/CTL	0,139	0,002	0,006	0,116	0,311
13.	02/27-18	02	27-18	F/CTL R AIL FAULT	0,139	0,002	0,006	0,116	0,311
14.	06/27-04	06	27-04	F/CTL ELAC 1 FAULT	0,133	0,002	0,006	0,116	0,304
15.	02/27-20	02	27-20	F/CTL SEC 2 FAULT	0,118	0,002	0,006	0,116	0,284
16.	02/27-17	02	27-17	F/CTL L R ELEV FAULT	0,132	0,002	0,007	0,048	0,215
17.	05/27-10	05	27-10	F/CTL FCDC 2 FAULT	0,104	0,001	0,004	0,056	0,184
18.	06/27-08	06	27-08	F/CTL ELEV SERVO FAULT	0,105	0,002	0,007	0,048	0,180
19.	06/27-30	06	27-30	SFCS	0,101	0,003	0,025	0,033	0,177
20.	02/27-30	02	27-30	SFCS	0,087	0,001	0,011	0,023	0,127
21.	06/27-15	06	27-15	F/CTL FLAPS LOCKED	0,083	0,001	0,003	0,031	0,121
22.	06/27-05	06	27-05	F/CTL ELAC 1 PITCH FAULT	0,083	0,001	0,003	0,031	0,121
23.	06/27-12	06	27-12	F/CTL FLAP SYS 1 FAULT	0,026	0,004	0,012	0,059	0,099

ATA 27- rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
24.	05/27-28	05	27-28	F/CTL	0,066	0,001	0,006	0,019	0,088
25.	02/27-12	02	27-12	F/CTL FLAP SYS 1 FAULT	0,067	0,001	0,002	0,019	0,083
26.	02/27-27	02	27-27	F/CTL SPLR FAULT	0,060	0,000	0,005	0,018	0,077
27.	02/27-04	02	27-04	F/CTL ELAC 1 FAULT	0,062	0,001	0,003	0,019	0,076
28.	05/27-30	05	27-30	SFCS	0,051	0,001	0,009	0,022	0,075
29.	02/27-28	02	27-28	F/CTL	0,057	0,000	0,003	0,016	0,066
30.	02/27-08	02	27-08	F/CTL ELEV SERVO FAULT	0,039	0,001	0,003	0,031	0,064
31.	04/27-30	04	27-30	SFCS	0,049	0,001	0,006	0,019	0,063
32.	06/27-02	06	27-02	F/CTL ALTN LAW	0,050	0,000	0,004	0,017	0,059
33.	09/27-30	09	27-30	SFCS	0,042	0,000	0,003	0,015	0,045
34.	03/27-30	03	27-30	SFCS	0,041	0,000	0,003	0,015	0,044
35.	08/27-30	08	27-30	SFCS	0,039	0,000	0,002	0,015	0,040
36.	07/27-30	07	27-30	SFCS	0,039	0,000	0,002	0,015	0,040
37.	02/27-05	02	27-05	F/CTL ELAC 1 PITCH FAULT	0,037	0,000	0,003	0,016	0,039
38.	06/27-13	06	27-13	F/CTL FLAP SYS 2 FAULT	0,033	0,001	0,002	0,019	0,038
39.	08/27-28	08	27-28	F/CTL	0,038	0,000	0,002	0,014	0,038
40.	06/27-03	06	27-03	F/CTL DIRECT LAW	0,038	0,000	0,002	0,014	0,038
41.	06/27-26	06	27-26	F/CTL SPD BRK DISAGREE	0,030	0,000	0,002	0,014	0,026
42.	06/27-19	06	27-19	F/CTL SEC 1 FAULT	0,031	0,000	0,001	0,013	0,026
43.	09/27-28	09	27-28	F/CTL	0,031	0,000	0,001	0,013	0,026
44.	08/27-22	08	27-22	F/CTL SIDESTICK PRIORITY	0,033	0,000	0,001	0,010	0,025
45.	04/27-09	04	27-09	F/CTL FCDC 1 FAULT	0,029	0,000	0,001	0,013	0,024
46.	02/27-10	02	27-10	F/CTL FCDC 2 FAULT	0,027	0,000	0,001	0,013	0,019
47.	06/27-14	06	27-14	F/CTL FLAPS FAULT	0,027	0,000	0,001	0,013	0,019

ATA 27- rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
49.	04/27-28	04	27-28	F/CTL	0,027	0,000	0,001	0,013	0,019
48.	02/27-23	02	27-23	F/CTL SLAT SYS 1 FAULT	0,027	0,000	0,001	0,013	0,019
50.	06/27-07	06	27-07	F/CTL ELAC 2 PITCH FAULT	0,027	0,000	0,001	0,013	0,019
51.	06/27-21	06	27-21	F/CTL SEC 3 FAULT	0,027	0,000	0,001	0,013	0,019
52.	09/27-12	09	27-12	F/CTL FLAP SYS 1 FAULT	0,027	0,000	0,001	0,013	0,019
53.	07/27-07	07	27-07	F/CTL ELAC 2 PITCH FAULT	0,022	0,000	0,001	0,012	0,011
54.	09/27-26	09	27-26	F/CTL SPD BRK DISAGREE	0,022	0,000	0,001	0,012	0,011
55.	07/27-01	07	27-01	F/CTL AIL SERVO FAULT	0,022	0,000	0,001	0,012	0,011
56.	08/27-09	08	27-09	F/CTL FCDC 1 FAULT	0,022	0,000	0,001	0,012	0,011
57.	08/27-02	08	27-02	F/CTL ALTN LAW	0,022	0,000	0,001	0,012	0,011
58.	06/27-25	06	27-25	F/CTL SLATS FAULT	0,022	0,000	0,001	0,012	0,011
59.	02/27-26	02	27-26	F/CTL SPD BRK DISAGREE	0,021	0,000	0,001	0,012	0,010
60.	02/27-11	02	27-11	F/CTL FLAP ATTACH SENSOR	0,021	0,000	0,001	0,012	0,010
62.	09/27-20	09	27-20	F/CTL SEC 2 FAULT	0,015	0,000	0,000	0,010	0,001
61.	05/27-07	05	27-07	F/CTL ELAC 2 PITCH FAULT	0,015	0,000	0,000	0,010	0,001
63.	02/27-14	02	27-14	F/CTL FLAPS FAULT	0,015	0,000	0,000	0,010	0,001
64.	02/27-25	02	27-25	F/CTL SLATS FAULT	0,015	0,000	0,000	0,010	0,001
65.	04/27-10	04	27-10	F/CTL FCDC 2 FAULT	0,015	0,000	0,000	0,010	0,001
66.	05/27-06	05	27-06	F/CTL ELAC 2 FAULT	0,015	0,000	0,000	0,010	0,001
67.	05/27-11	05	27-11	F/CTL FLAP ATTACH SENSOR	0,015	0,000	0,000	0,010	0,001
68.	05/27-13	05	27-13	F/CTL FLAP SYS 2 FAULT	0,015	0,000	0,000	0,010	0,001
70.	02/27-22	02	27-22	F/CTL SIDESTICK PRIORITY	0,015	0,000	0,000	0,010	0,001
71.	04/27-01	04	27-01	F/CTL AIL SERVO FAULT	0,015	0,000	0,000	0,010	0,001
79.	07/27-19	07	27-19	F/CTL SEC 1 FAULT	0,015	0,000	0,000	0,010	0,001

ATA 27- rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
81.	08/27-01	08	27-01	F/CTL AIL SERVO FAULT	0,015	0,000	0,000	0,010	0,001
83.	08/27-26	08	27-26	F/CTL SPD BRK DISAGREE	0,015	0,000	0,000	0,010	0,001
69.	02/27-13	02	27-13	F/CTL FLAP SYS 2 FAULT	0,015	0,000	0,000	0,010	0,001
72.	04/27-02	04	27-02	F/CTL ALTN LAW	0,015	0,000	0,000	0,010	0,001
73.	04/27-05	04	27-05	F/CTL ELAC 1 PITCH FAULT	0,015	0,000	0,000	0,010	0,001
74.	04/27-07	04	27-07	F/CTL ELAC 2 PITCH FAULT	0,015	0,000	0,000	0,010	0,001
75.	05/27-01	05	27-01	F/CTL AIL SERVO FAULT	0,015	0,000	0,000	0,010	0,001
76.	05/27-05	05	27-05	F/CTL ELAC 1 PITCH FAULT	0,015	0,000	0,000	0,010	0,001
77.	06/27-11	06	27-11	F/CTL FLAP ATTACH SENSOR	0,015	0,000	0,000	0,010	0,001
78.	06/27-20	06	27-20	F/CTL SEC 2 FAULT	0,015	0,000	0,000	0,010	0,001
80.	07/27-26	07	27-26	F/CTL SPD BRK DISAGREE	0,015	0,000	0,000	0,010	0,001
82.	08/27-20	08	27-20	F/CTL SEC 2 FAULT	0,015	0,000	0,000	0,010	0,001
84.	09/27-02	09	27-02	F/CTL ALTN LAW	0,015	0,000	0,000	0,010	0,001
85.	03/27-28	03	27-28	F/CTL	0,015	0,000	0,000	0,010	0,001
87.	05/27-26	05	27-26	F/CTL SPD BRK DISAGREE	0,015	0,000	0,000	0,010	0,001
88.	09/27-10	09	27-10	F/CTL FCDC 2 FAULT	0,015	0,000	0,000	0,010	0,001
86.	02/27-24	02	27-24	F/CTL SLAT SYS 2 FAULT	0,015	0,000	0,000	0,010	0,001
89.	05/27-27	05	27-27	F/CTL SPLR FAULT	0,015	0,000	0,000	0,010	0,001
90.	02/27-29	02	27-29	F/CTPLR FAULT	0,015	0,000	0,000	0,010	0,001
91.	09/27-08	09	27-08	F/CTL ELEV SERVO FAULT	0,015	0,000	0,000	0,010	0,001
92.	08/27-10	08	27-10	F/CTL FCDC 2 FAULT	0,015	0,000	0,000	0,010	0,001
93.	09/27-06	09	27-06	F/CTL ELAC 2 FAULT	0,015	0,000	0,000	0,010	0,001
94.	06/27-28	06	27-28	F/CTL	0,013	0,000	0,001	0,012	0,000

Prilog D.3. Rang značajki za sustav ATA 29

ATA 29 – rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
1.	06/29-15	06	29-15	HYD Y SYS LO PR	0,3350	0,0411	0,0736	0,3400	1,0000
2.	06/29-12	06	29-12	HYD Y ENG 2 PUMP LO PR	0,3350	0,0411	0,0736	0,3400	1,0000
3.	08/29-13	08	29-13	HYD Y RSVR LO LVLXX	0,3350	0,0411	0,0736	0,3400	1,0000
4.	09/29-15	09	29-15	HYD Y SYS LO PR	0,3350	0,0411	0,0736	0,3400	1,0000
5.	06/29-13	06	29-13	HYD Y RSVR LO LVLXX	0,3350	0,0411	0,0736	0,3400	1,0000
6.	02/29-9	02	29-9	HYD PTU FAULT	0,3730	0,0510	0,0930	0,1095	0,7932
7.	02/29-12	02	29-12	HYD Y ENG 2 PUMP LO PR	0,2746	0,0304	0,0942	0,1281	0,6677
8.	02/29-4	02	29-4	HYD G ENG 1 PUMP LO PR	0,3183	0,0371	0,0692	0,0774	0,6357
9.	02/29-6	02	29-6	HYD G SYS LO PR	0,3183	0,0371	0,0692	0,0774	0,6357
10.	02/29-8	02	29-8	HYD GXY SYS LO PRXX	0,2256	0,0186	0,0599	0,1034	0,5161
11.	06/29-1	06	29-1	HYD B RSVR LO AIR PRXX	0,1916	0,0134	0,0439	0,0915	0,4311
12.	02/29-3	02	29-3	HYD BXY SYS LO PRXX	0,1535	0,0086	0,0286	0,0791	0,3418
13.	02/29-7	02	29-7	HYD GXB SYS LO PRX	0,1535	0,0086	0,0286	0,0791	0,3418
14.	02/29-2	02	29-2	HYD B SYS LO PR	0,0998	0,0214	0,0370	0,0791	0,3005
15.	02/29-0	02	29-0	HYD B ELEC PUMP LO PRX	0,1066	0,0042	0,0140	0,0648	0,2401
16.	06/29-6	06	29-6	HYD G SYS LO PR	0,1066	0,0042	0,0140	0,0648	0,2401
17.	02/29-10	02	29-10	HYD RAT FAULT	0,1066	0,0042	0,0140	0,0648	0,2401
18.	05/29-4	05	29-4	HYD G ENG 1 PUMP LO PR	0,1066	0,0042	0,0140	0,0648	0,2401
19.	05/29-9	05	29-9	HYD PTU FAULT	0,1066	0,0042	0,0140	0,0648	0,2401
20.	06/29-2	06	29-2	HYD B SYS LO PR	0,1066	0,0042	0,0140	0,0648	0,2401
21.	08/29-6	08	29-6	HYD G SYS LO PR	0,1066	0,0042	0,0140	0,0648	0,2401
22.	09/29-0	09	29-0	HYD B ELEC PUMP LO PRX	0,1066	0,0042	0,0140	0,0648	0,2401
23.	06/29-4	06	29-4	HYD G ENG 1 PUMP LO PR	0,1066	0,0042	0,0140	0,0648	0,2401

ATA 29 – rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
24.	06/29-5	06	29-5	HYD G RSVR LO LVLXX	0,1066	0,0042	0,0140	0,0648	0,2401
25.	07/29-6	07	29-6	HYD G SYS LO PR	0,1066	0,0042	0,0140	0,0648	0,2401
26.	09/29-2	09	29-2	HYD B SYS LO PR	0,1066	0,0042	0,0140	0,0648	0,2401
27.	02/29-11	02	29-11	HYD Y ELEC PUMP OVHTXX	0,0000	0,0000	0,0000	0,0000	0,0000
28.	06/29-11	06	29-11	HYD Y ELEC PUMP OVHTXX	0,0000	0,0000	0,0000	0,0000	0,0000
29.	06/29-14	06	29-14	HYD Y RSVR OVHTX	0,0000	0,0000	0,0000	0,0000	0,0000
30.	06/29-16	06	29-16	HYD B RSVR LO LVLXX	0,0000	0,0000	0,0000	0,0000	0,0000
31.	06/29-8	06	29-8	HYD GXY SYS LO PRXX	0,0000	0,0000	0,0000	0,0000	0,0000
32.	09/29-13	09	29-13	HYD Y RSVR LO LVLXX	0,0000	0,0000	0,0000	0,0000	0,0000
33.	02/29-15	02	29-15	HYD Y SYS LO PR	0,0000	0,0000	0,0000	0,0000	0,0000

Prilog D.4. Rang značajki za sustav ATA 32

ATA 32 – rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
1.	02/32-6	02	32-6	BRAKES SYS 2 FAULT	0,3464	0,0634	0,1139	0,3847	1,0000
2.	08/32-1	08	32-1	BRAKES AUTO BRK FAULT	0,2504	0,0230	0,0411	0,2949	0,6707
3.	07/32-6	07	32-6	BRAKES SYS 2 FAULT	0,3129	0,0359	0,0583	0,1805	0,6468
4.	03/32-1	03	32-1	BRAKES AUTO BRK FAULT	0,2083	0,0230	0,0411	0,2949	0,6243
5.	06/32-10	06	32-10	L/G LGCIU 1 FAULT	0,2038	0,0152	0,0272	0,2702	0,5684
6.	08/32-12	08	32-12	L/G SYS DISAGREE	0,2015	0,0139	0,0223	0,2372	0,5228
7.	06/32-1	06	32-1	BRAKES AUTO BRK FAULT	0,1436	0,0076	0,0135	0,2372	0,4424
8.	09/32-14	09	32-14	WHEEL N/W STRG FAULT	0,1436	0,0076	0,0135	0,2372	0,4424
9.	02/32-12	02	32-12	L/G SYS DISAGREE	0,1436	0,0076	0,0135	0,2372	0,4424
10.	05/32-6	05	32-6	BRAKES SYS 2 FAULT	0,0817	0,0152	0,0272	0,2702	0,4341
11.	06/32-4	06	32-4	BRAKES HOT	0,0692	0,0152	0,0272	0,2702	0,4202
12.	02/32-4	02	32-4	BRAKES HOT	0,1274	0,0100	0,0328	0,0828	0,2785
13.	02/32-14	02	32-14	WHEEL N/W STRG FAULT	0,0356	0,0214	0,0346	0,1292	0,2430
14.	06/32-2	06	32-2	BRAKES BSCU CH 1 FAULT	0,1016	0,0056	0,0188	0,0700	0,2158
15.	02/32-7	02	32-7	L/G DOORS NOT CLOSED	0,1013	0,0040	0,0133	0,0641	0,2010
16.	05/32-4	05	32-4	BRAKES HOT	0,0981	0,0040	0,0133	0,0641	0,1975
17.	02/32-2	02	32-2	BRAKES BSCU CH 1 FAULT	0,0925	0,0031	0,0106	0,0607	0,1838
18.	04/32-4	04	32-4	BRAKES HOT	0,0876	0,0031	0,0106	0,0607	0,1783
19.	06/32-6	06	32-6	BRAKES SYS 2 FAULT	0,0873	0,0031	0,0106	0,0607	0,1781
20.	05/32-7	05	32-7	L/G DOORS NOT CLOSED	0,0873	0,0031	0,0106	0,0607	0,1781
21.	09/32-4	09	32-4	BRAKES HOT	0,0809	0,0031	0,0106	0,0607	0,1709
22.	08/32-4	08	32-4	BRAKES HOT	0,0809	0,0031	0,0106	0,0607	0,1709
23.	07/32-5	07	32-5	BRAKES SYS 1 FAULT	0,0799	0,0023	0,0079	0,0569	0,1618

ATA 32 – rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
24.	06/32-11	06	32-11	L/G LGCIU 2 FAULTXX	0,0799	0,0023	0,0079	0,0569	0,1618
25.	06/32-3	06	32-3	BRAKES BSCU CH 2 FAULT	0,0786	0,0023	0,0079	0,0569	0,1605
26.	03/32-4	03	32-4	BRAKES HOT	0,0701	0,0023	0,0079	0,0569	0,1511
27.	08/32-6	08	32-6	BRAKES SYS 2 FAULT	0,0650	0,0015	0,0053	0,0523	0,1366
28.	04/32-1	04	32-1	BRAKES AUTO BRK FAULT	0,0650	0,0015	0,0053	0,0523	0,1366
29.	05/32-10	05	32-10	L/G LGCIU 1 FAULT	0,0650	0,0015	0,0053	0,0523	0,1366
30.	09/32-1	09	32-1	BRAKES AUTO BRK FAULT	0,0650	0,0015	0,0053	0,0523	0,1366
31.	06/32-8	06	32-8	L/G GEAR NOT DOWNLOCKEDX	0,0650	0,0015	0,0053	0,0523	0,1366
32.	07/32-4	07	32-4	BRAKES HOT	0,0581	0,0015	0,0053	0,0523	0,1290
33.	02/32-5	02	32-5	BRAKES SYS 1 FAULT	0,0447	0,0015	0,0053	0,0523	0,1143
34.	05/32-1	05	32-1	BRAKES AUTO BRK FAULT	0,0458	0,0008	0,0026	0,0461	0,1049
35.	09/32-10	09	32-10	L/G LGCIU 1 FAULT	0,0458	0,0008	0,0026	0,0461	0,1049
36.	09/32-3	09	32-3	BRAKES BSCU CH 2 FAULT	0,0458	0,0008	0,0026	0,0461	0,1049
37.	02/32-10	02	32-10	L/G LGCIU 1 FAULT	0,0458	0,0008	0,0026	0,0461	0,1049
38.	05/32-11	05	32-11	L/G LGCIU 2 FAULTXX	0,0458	0,0008	0,0026	0,0461	0,1049
39.	08/32-5	08	32-5	BRAKES SYS 1 FAULT	0,0458	0,0008	0,0026	0,0461	0,1049
40.	09/32-5	09	32-5	BRAKES SYS 1 FAULT	0,0458	0,0008	0,0026	0,0461	0,1049
41.	02/32-3	02	32-3	BRAKES BSCU CH 2 FAULT	0,0458	0,0008	0,0026	0,0461	0,1049
42.	09/32-2	09	32-2	BRAKES BSCU CH 1 FAULT	0,0458	0,0008	0,0026	0,0461	0,1049
43.	07/32-2	07	32-2	BRAKES BSCU CH 1 FAULT	0,0306	0,0008	0,0026	0,0461	0,0881
44.	02/32-1	02	32-1	BRAKES AUTO BRK FAULT	0,0133	0,0015	0,0053	0,0523	0,0797
45.	06/32-5	06	32-5	BRAKES SYS 1 FAULT	0,0065	0,0036	0,0060	0,0523	0,0753
46.	09/32-6	09	32-6	BRAKES SYS 2 FAULT	0,0302	0,0003	0,0006	0,0044	0,0392
47.	06/32-7	06	32-7	L/G DOORS NOT CLOSED	0,0302	0,0003	0,0006	0,0044	0,0392
48.	02/32-0	02	32-0	BRAKES A/SKID N/WS FAULT	0,0180	0,0001	0,0002	0,0012	0,0214



ATA 32 – rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
49.	02/32-13	02	32-13	WHEEL N.W.STEER FAULTX	0,0000	0,0000	0,0000	0,0000	0,0000
50.	03/32-5	03	32-5	BRAKES SYS 1 FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
51.	03/32-6	03	32-6	BRAKES SYS 2 FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
52.	05/32-5	05	32-5	BRAKES SYS 1 FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
53.	06/32-0	06	32-0	BRAKES A/SKID N/WS FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
54.	06/32-14	06	32-14	WHEEL N/W STRG FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
55.	07/32-1	07	32-1	BRAKES AUTO BRK FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
56.	07/32-14	07	32-14	WHEEL N/W STRG FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
57.	05/32-9	05	32-9	L/G GEAR NOT UNLOCKEDX	0,0000	0,0000	0,0000	0,0000	0,0000
58.	06/32-9	06	32-9	L/G GEAR NOT UNLOCKEDX	0,0000	0,0000	0,0000	0,0000	0,0000
59.	07/32-3	07	32-3	BRAKES BSCU CH 2 FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
60.	04/32-6	04	32-6	BRAKES SYS 2 FAULT	0,0000	0,0000	0,0000	0,0000	0,0000

Prilog D.5. Rang značajki za sustav ATA 34

ATA 34 – rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
1.	03/34-19	03	34-19	NAV ILS 2 FAULT	0,2676	0,0098	0,0217	0,4279	1,0000
2.	04/34-23	04	34-23	IR	0,2475	0,0098	0,0217	0,4279	0,9724
3.	08/34-24	08	34-24	NAV PRED W/S DET FAULT	0,1890	0,0049	0,0108	0,3818	0,8066
4.	08/34-17	08	34-17	NAV ILS 1 FAULT	0,1890	0,0049	0,0108	0,3818	0,8066
5.	09/34-11	09	34-11	NAV GPS1 FAULT	0,1890	0,0049	0,0108	0,3818	0,8066
6.	04/34-6	04	34-6	NAV ADR 3 FAULT	0,2036	0,0106	0,0239	0,2646	0,6913
7.	06/34-10	06	34-10	NAV FM/GPS POS DISAGREE	0,0962	0,0049	0,0108	0,3818	0,6791
8.	06/34-22	06	34-22	NAV IR 3 FAULTXX	0,1597	0,0142	0,0345	0,2196	0,5887
9.	02/34-3	02	34-3	NAV ADR 1X3 FAULTXX	0,1698	0,0063	0,0165	0,1323	0,4470
10.	05/34-11	05	34-11	NAV GPS1 FAULT	0,1231	0,0021	0,0054	0,1060	0,3253
11.	02/34-19	02	34-19	NAV ILS 2 FAULT	0,1305	0,0023	0,0072	0,0574	0,2715
12.	03/34-10	03	34-10	NAV FM/GPS POS DISAGREE	0,0919	0,0012	0,0035	0,0497	0,2011
13.	06/34-19	06	34-19	NAV ILS 2 FAULT	0,0739	0,0023	0,0072	0,0574	0,1936
14.	06/34-3	06	34-3	NAV ADR 1X3 FAULTXX	0,1025	0,0014	0,0050	0,0316	0,1933
15.	05/34-24	05	34-24	NAV PRED W/S DET FAULT	0,0729	0,0010	0,0093	0,0231	0,1462
16.	06/34-20	06	34-20	NAV IR 1 FAULT	0,0720	0,0007	0,0024	0,0271	0,1407
17.	02/34-17	02	34-17	NAV ILS 1 FAULT	0,0429	0,0012	0,0035	0,0497	0,1337
18.	06/34-24	06	34-24	NAV PRED W/S DET FAULT	0,0619	0,0008	0,0076	0,0217	0,1265
19.	06/34-6	06	34-6	NAV ADR 3 FAULT	0,0579	0,0010	0,0093	0,0231	0,1255
20.	06/34-14	06	34-14	NAV GPWS TERR DET FAULT	0,0487	0,0007	0,0024	0,0271	0,1086
21.	02/34-6	02	34-6	NAV ADR 3 FAULT	0,0530	0,0006	0,0056	0,0198	0,1086
22.	06/34-12	06	34-12	NAV GPS2 FAULTXX	0,0329	0,0014	0,0050	0,0316	0,0976
23.	04/34-10	04	34-10	NAV FM/GPS POS DISAGREE	0,0486	0,0003	0,0032	0,0173	0,0954

ATA 34 – rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
24.	06/34-27	06	34-27	NAV TCAS FAULT	0,0424	0,0006	0,0056	0,0198	0,0939
25.	02/34-13	02	34-13	NAV GPWS FAULT	0,0134	0,0012	0,0035	0,0497	0,0932
26.	02/34-23	02	34-23	IR	0,0445	0,0014	0,0056	0,0160	0,0929
27.	02/34-10	02	34-10	NAV FM/GPS POS DISAGREE	0,0437	0,0005	0,0045	0,0188	0,0928
28.	02/34-27	02	34-27	NAV TCAS FAULT	0,0434	0,0005	0,0045	0,0188	0,0925
29.	06/34-13	06	34-13	NAV GPWS FAULT	0,0386	0,0004	0,0036	0,0177	0,0827
30.	04/34-24	04	34-24	NAV PRED W/S DET FAULT	0,0396	0,0003	0,0029	0,0168	0,0821
31.	02/34-12	02	34-12	NAV GPS2 FAULTXX	0,0387	0,0002	0,0023	0,0159	0,0785
32.	06/34-11	06	34-11	NAV GPS1 FAULT	0,0358	0,0003	0,0026	0,0164	0,0756
33.	05/34-27	05	34-27	NAV TCAS FAULT	0,0350	0,0002	0,0023	0,0159	0,0734
34.	06/34-17	06	34-17	NAV ILS 1 FAULT	0,0332	0,0002	0,0023	0,0159	0,0709
35.	07/34-14	07	34-14	NAV GPWS TERR DET FAULT	0,0349	0,0001	0,0010	0,0135	0,0681
36.	02/34-26	02	34-26	NAV RA 2 FAULT	0,0322	0,0002	0,0016	0,0148	0,0671
37.	03/34-24	03	34-24	NAV PRED W/S DET FAULT	0,0305	0,0002	0,0016	0,0148	0,0647
38.	09/34-6	09	34-6	NAV ADR 3 FAULT	0,0305	0,0001	0,0013	0,0142	0,0634
39.	07/34-24	07	34-24	NAV PRED W/S DET FAULT	0,0301	0,0001	0,0013	0,0142	0,0629
40.	06/34-16	06	34-16	NAV IAS DISCREPANCY	0,0145	0,0007	0,0024	0,0271	0,0615
41.	04/34-22	04	34-22	NAV IR 3 FAULTXX	0,0288	0,0001	0,0013	0,0142	0,0610
42.	08/34-14	08	34-14	NAV GPWS TERR DET FAULT	0,0263	0,0001	0,0010	0,0135	0,0562
43.	05/34-23	05	34-23	IR	0,0222	0,0005	0,0017	0,0158	0,0554
44.	02/34-24	02	34-24	NAV PRED W/S DET FAULT	0,0172	0,0009	0,0035	0,0187	0,0554
45.	06/34-9	06	34-9	NAV ATT DISCREPANCY	0,0248	0,0001	0,0010	0,0135	0,0541
46.	06/34-1	06	34-1	NAV ADR 1 FAULT	0,0248	0,0001	0,0010	0,0135	0,0541
47.	07/34-10	07	34-10	NAV FM/GPS POS DISAGREE	0,0215	0,0001	0,0006	0,0126	0,0478
48.	09/34-14	09	34-14	NAV GPWS TERR DET FAULT	0,0215	0,0001	0,0006	0,0126	0,0478

ATA 34 – rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
49.	04/34-27	04	34-27	NAV TCAS FAULT	0,0215	0,0001	0,0006	0,0126	0,0478
50.	02/34-25	02	34-25	NAV RA 1 FAULTXX	0,0215	0,0001	0,0006	0,0126	0,0478
51.	06/34-0	06	34-0	ADRX	0,0215	0,0001	0,0006	0,0126	0,0478
52.	09/34-24	09	34-24	NAV PRED W/S DET FAULT	0,0215	0,0001	0,0006	0,0126	0,0478
53.	07/34-23	07	34-23	IR	0,0215	0,0001	0,0006	0,0126	0,0478
54.	05/34-13	05	34-13	NAV GPWS FAULT	0,0215	0,0001	0,0006	0,0126	0,0478
55.	09/34-13	09	34-13	NAV GPWS FAULT	0,0215	0,0001	0,0006	0,0126	0,0478
56.	06/34-21	06	34-21	NAV IR 2 FAULTXX	0,0215	0,0001	0,0006	0,0126	0,0478
57.	09/34-10	09	34-10	NAV FM/GPS POS DISAGREE	0,0215	0,0001	0,0006	0,0126	0,0478
58.	09/34-27	09	34-27	NAV TCAS FAULT	0,0215	0,0001	0,0006	0,0126	0,0478
59.	08/34-10	08	34-10	NAV FM/GPS POS DISAGREE	0,0204	0,0001	0,0006	0,0126	0,0463
60.	07/34-6	07	34-6	NAV ADR 3 FAULT	0,0204	0,0001	0,0006	0,0126	0,0463
61.	02/34-11	02	34-11	NAV GPS1 FAULT	0,0296	0,0001	0,0005	0,0034	0,0463
62.	03/34-27	03	34-27	NAV TCAS FAULT	0,0152	0,0000	0,0003	0,0113	0,0368
63.	06/34-18	06	34-18	NAV ILS 1X2 FAULTXX	0,0152	0,0000	0,0003	0,0113	0,0368
64.	05/34-6	05	34-6	NAV ADR 3 FAULT	0,0152	0,0000	0,0003	0,0113	0,0368
65.	09/34-23	09	34-23	IR	0,0152	0,0000	0,0003	0,0113	0,0368
66.	02/34-22	02	34-22	NAV IR 3 FAULTXX	0,0152	0,0000	0,0003	0,0113	0,0368
67.	05/34-0	05	34-0	ADRX	0,0152	0,0000	0,0003	0,0113	0,0368
68.	04/34-12	04	34-12	NAV GPS2 FAULTXX	0,0152	0,0000	0,0003	0,0113	0,0368
69.	06/34-15	06	34-15	NAV HDG DISCREPANCY	0,0152	0,0000	0,0003	0,0113	0,0368
70.	03/34-6	03	34-6	NAV ADR 3 FAULT	0,0152	0,0000	0,0003	0,0113	0,0368
71.	08/34-6	08	34-6	NAV ADR 3 FAULT	0,0152	0,0000	0,0003	0,0113	0,0368
72.	09/34-3	09	34-3	NAV ADR 1X3 FAULTXX	0,0152	0,0000	0,0003	0,0113	0,0368
73.	02/34-2	02	34-2	NAV ADR 1X2 FAULTXX	0,0152	0,0000	0,0003	0,0113	0,0368

ATA 34 – rang značajki									
R. br.	Značajka	Faza leta	Šifrirana WN poruka	WN poruka	KP	GI	IG	IGR	Rang (ukupni)
74.	02/34-5	02	34-5	NAV ADR 2X3 FAULTXX	0,0152	0,0000	0,0003	0,0113	0,0368
75.	05/34-12	05	34-12	NAV GPS2 FAULTXX	0,0152	0,0000	0,0003	0,0113	0,0368
76.	06/34-23	06	34-23	IR	0,0033	0,0003	0,0032	0,0173	0,0332
77.	02/34-14	02	34-14	NAV GPWS TERR DET FAULT	0,0021	0,0003	0,0026	0,0164	0,0294
78.	06/34-8	06	34-8	NAV ALTI DISCREPANCY	0,0023	0,0001	0,0013	0,0142	0,0247
79.	03/34-12	03	34-12	NAV GPS2 FAULTXX	0,0000	0,0000	0,0000	0,0000	0,0000
80.	03/34-23	03	34-23	IR	0,0000	0,0000	0,0000	0,0000	0,0000
81.	06/34-7	06	34-7	NAV ADR DISAGREE	0,0000	0,0000	0,0000	0,0000	0,0000
82.	07/34-27	07	34-27	NAV TCAS FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
83.	08/34-13	08	34-13	NAV GPWS FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
84.	09/34-12	09	34-12	NAV GPS2 FAULTXX	0,0000	0,0000	0,0000	0,0000	0,0000
85.	09/34-22	09	34-22	NAV IR 3 FAULTXX	0,0000	0,0000	0,0000	0,0000	0,0000
86.	07/34-11	07	34-11	NAV GPS1 FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
87.	07/34-13	07	34-13	NAV GPWS FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
88.	08/34-11	08	34-11	NAV GPS1 FAULT	0,0000	0,0000	0,0000	0,0000	0,0000
89.	08/34-23	08	34-23	IR	0,0000	0,0000	0,0000	0,0000	0,0000
90.	10/34-27	10	34-27	NAV TCAS FAULT	0,0000	0,0000	0,0000	0,0000	0,0000

## ŽIVOTOPIS

Marina Tošić (rođ. 1987.) završila je Opću gimnaziju u Zlataru 2005. godine, nakon čega upisuje Fakultet strojarstva i brodogradnje, Sveučilišta u Zagrebu. Godine 2011. diplomirala je s temom rada „Prijedlog unapređenja upravljanja infrastrukturnim objektima“. Iste godine upisuje poslijediplomski doktorski studij na Fakultetu strojarstva i brodogradnje. Prvo radno iskustvo stječe u tvrtki Končar D&ST kao pripravnik na projektu izgradnje visokonaponskog laboratorija. Od veljače 2012. godine zaposlena je na radnom mjestu asistenta na Zavodu za industrijsko inženjerstvo. Uključena je u nastavne aktivnosti Zavoda na kolegijima „Održavanje“, „Menadžment u održavanju“, „Ekologija i održavanje“ i „Informacijski sustavi“ i asistira u vođenju završnih i diplomskih radova. Sudjeluje u organizacijskim aktivnostima Zavodima te je članica Povjerenstva za e-učenje na Fakultetu strojarstva i brodogradnje. Autorica ili koautorica je osam znanstvenih radova. Područja njezina istraživačkog interesa uključuju održavanje, dubinsku analizu podataka i metode višekriterijalnog odlučivanja i planiranja. Aktivno se služi engleskim i španjolskim jezikom u govoru i pismu, a pasivno njemačkim.

## **CURRICULUM VITAE**

Marina Tošić (born 1987) has finished high school in 2005, after which she enrolled in the Faculty of Mechanical Engineering and Naval Architecture (UNIZAG FSB), University of Zagreb. In 2011, she graduated with the thesis entitled: “Proposal of Facility Management Upgrading” and enrolled in the Postgraduate course at UNIZAG FSB in the same year. She gained her first working experience in the Končar D&ST Company as an intern on a construction project of a High Volt Laboratory. From February 2012 she is employed at the Department of Industrial Engineering at the UNIZAG FSB as a research and teaching assistant. She is involved in Department’s educational activities within the courses on “Maintenance”, “Maintenance Management”, “Ecology and Maintenance” and “Information Systems”, where she is also assisting in supervising bachelor and master thesis. She participates in Department’s organizational activities and is a member of the E-learning Committee at UNIZAG FSB. She is author or co-author of 8 scientific papers. Her research interests include maintenance, data mining techniques and multi-criteria decision-making techniques. She is fluent in English and Spanish and speaks some German.